

INDIVIDUAL DIFFERENCES IN INTENTIONAL AND INCIDENTAL LEARNING

BY MAVIS PLENDERLEITH AND LEO POSTMAN

Department of Psychology, University of California

This study investigates the role of linguistic habits in the intentional and incidental learning of verbal materials. Linguistic habits determine the differential responses made to the stimulus materials during practice. The stronger the differential responses evoked by the stimulus items the better they are recalled. Similarly, individual differences in amount of recall reflect variations in the strength (conventionality) of differential responses. These relationships are more pronounced when learning is incidental than when it is intentional, and also are stronger with nonsense than with meaningful learning materials.

I. INTRODUCTION

Activity is a pervasive characteristic of verbal learning. Whenever the task exceeds the subject's immediate memory span, only a fraction of the material can be fixated on any one trial. The order in which different parts of the material are retained is not random but reflects the 'complex and subtle influence exercised by the behavior repertoire of the individual upon the fixation of new responses' (McGeoch, 1942, p. 304). Specifically, the transfer of earlier habits to a new task determines the subject's differential responses to the learning materials.

The importance of differential responses as mediators of learning has recently been stressed by several writers (Dollard & Miller, 1950; Mandler, 1954; Osgood, 1953). Differential responses are reactions which give identity and meaning to a stimulus. When stimuli are verbal, differential responses are likely to be themselves verbal or symbolic. These responses may be explicit, for example, overt verbal associations, or implicit. Implicit responses may be conceived as fractional components of responses which originally had an explicit (cf. Osgood, 1953, pp. 397f.). A well-established implicit response may carry the meaning of a stimulus concurrently with other associations or in the absence of overt associations. In the memorization of verbal series, differential responses (whether explicit or implicit) may be assumed to play a dual part: (a) they are instrumental in the differentiation of the individual items, and (b) they serve to mediate the reproduction of the items at the time of the memory test. Since differential responses are well practised and familiar reactions, they are readily associated with the context of the learning situation and then can 'lead' the subject back to the learning items proper.

Verbal stimuli may vary widely with respect to the frequency with which they elicit differential responses in a sample of subjects (Glaze, 1928; Hull, 1933). The positive relationship between associative frequency (association value) and recall has long been recognized (McGeoch, 1942, pp. 157f.). Frequency of recall should also be a function of the associative potency of the learning items. The associative potency of an item refers to the strength of the differential responses which that item elicits (cf. Mandler, 1955). For a given verbal stimulus, degree of associative potency may be estimated from the amount of agreement in subjects' associative responses. The use of inter-subject agreement as a measure of associative potency is in accord with Marbe's Law, which states that the

latency of associative responses varies inversely with the frequency of responses in the population (Woodworth & Schlosberg, 1954, pp. 61-5). Since latency may be used as an index of habit strength, degree of inter-subject agreement provides us with an estimate of the relative strength of differential responses to a given word.

The responses of individual subjects show different degrees of conformity to population norms. We shall refer to this dimension of individual variation as the *conventionality* of associative responses. Applying Marbe's Law as before, we assume that the degree of a subject's conventionality provides an estimate of the strength of his associative responses. The more conventional a subject's explicit differential responses to the learning items the better should be his retention of these items. The effect of individual differences in conventionality should, however, vary with the nature of the materials and the set of the learner. With items of high associative potency, the explicit differential response given during practice may be of little consequence. Even if they are not made explicitly conventional responses are likely to occur implicitly and will serve to mediate the later reproduction of the items. With items of low associative potency, on the other hand, the differentiation and subsequent reproduction of the learning materials may depend largely on the explicit differential responses made during practice. Thus, the conventionality of such responses should be more important in the learning of nonsense material than of meaningful materials since the latter are apt to have a higher degree of associative potency.

As regards the set of the learner, the conventionality of explicit differential response should be a more critical variable when learning is incidental than when it is intentional. Intentional learners will maximize whatever differential responses are available to them for purposes of differentiating and connecting the learning items. The motivation to learn may mask, at least in part, the effects of individual differences in verbal habits. When subjects are not motivated to learn, only strong differential responses will be elicited. Individual differences in conventionality should, therefore, have pronounced effects on the amount of incidental learning. An earlier study (Plenderleith & Postmar, 1956) has already shown that incidental learning is more sensitive than intentional learning to individual differences in verbal and discriminative habits.

The present study tests the following hypotheses: (a) items of high associative potency will be recalled better than items of low associative potency; (b) the more conventional an individual's explicit differential responses the better will be his retention performance; (c) individual differences in conventionality will have more significant effects with items of low associative potency than with items of high associative potency; (d) individual differences in conventionality will have more significant effects under conditions of incidental than intentional learning.

II. THE EXPERIMENTAL PROCEDURE

(i) *Experimental design.* Two sets of learning materials—a series of nonsense syllables and a series of English words—were used. There were two groups of subjects, viz. one group of intentional learners and one group of incidental learners. Each group learned both sets of learning materials. Associative responses to the learning items were obtained during practice, and recall was tested immediately after practice by the method of free recall.

(ii) *Learning procedures and tests of association.* The first learning task consisted of a series of 20 nonsense syllables varying widely in association value according to Glaze's (1928) norms. There were five syllables each of 0, 33, 66, and 100% association value. One syllable of each of these association values

appeared in each of five blocks of four syllables in the list. The syllables, printed in capital letters, were presented for 4 sec. on 2×2 lantern slides. Following each exposure the subjects wrote down a meaningful association suggested by the syllable. The associations were written in the appropriately numbered space of a record blank. A cardboard strip was used to cover all spaces except the one in which the subjects were writing. The subjects were given 6 sec. in which to record their association, after which the next slide was exposed. The intentional learners were, in addition, instructed to remember as many of the syllables as possible without regard to the order of presentation. No learning instructions were given to the incidental subjects.

A 5 min. test of free recall was given immediately after the presentation of the last syllable. The subjects were instructed to write down all the syllables they could remember. They were encouraged to write down guesses as well as items of which they felt sure. Incidental learners were asked whether they had attempted to learn the syllables or had expected a test of recall. Five subjects who gave evidence of having guessed the purpose of the experiment were replaced.

The second task consisted of the 100 stimulus words of the Kent-Rosanoff test of word association. The stimulus words were printed in columns of 25 each on two pages of a test booklet. For each stimulus word, the subjects were given 6 sec. in which to write down the first association that came to mind. The associations were written in a space next to the stimulus word. A cardboard strip was used to cover all stimulus words except the one to which the subjects were responding. The intentional subjects, but not the incidental subjects, were instructed to remember as many of the stimulus words as possible without regard to the order of presentation. A 10 min. test of free recall was given immediately after the association test. At the end of the recall test, incidental subjects were asked whether they had expected the recall test or had attempted to memorize the words. Only two subjects had to be replaced on the basis of their answers to these questions. (The failure of most subjects to guess the purpose of the second task was probably due to the use of the Kent-Rosanoff test. The subjects recognized this test as a procedure for personality assessment and hence did not anticipate the second test of recall.)

III. RESULTS

(i) *The association tests.* To test our experimental hypotheses, it was necessary (a) to divide both nonsense syllables and words into items of high and low associative potency, and (b) to measure the conventionality of subjects' responses to these items.

The associative potency of a verbal stimulus has been defined in terms of the amount of agreement in subjects' associative responses to that stimulus. For both the nonsense syllables and the words, sets of norms were available from which the degree of inter-subject agreement could be determined. For the nonsense syllables, the associative responses of 530 undergraduate students had been obtained in earlier investigations at the University of California. The frequencies of different associations were tabulated for each of the syllables, and the frequency of the most popular association was used as a measure of inter-subject agreement. On the basis of these measures, the syllables were divided into two groups of 10 each, one of high associative potency, and the other of low associative potency. For the syllables of high associative potency, the mean percentage of subjects giving the most popular response is 44.34, with a standard deviation of 12.26. For the syllable of low associative potency, the mean percentage of subjects giving the most popular response is 17.00, with a standard deviation of 8.83. In the case of the words, the recent college norms of the Kent-Rosanoff test (Russell & Jenkins, 1954) were used. The frequency of the most popular response was again used to measure inter-subject agreement. The words were divided into two groups of 50 each, one of high and one of low associative potency. For the words of high associative potency the mean percentage of subjects giving the most popular response is 53.90, with a standard deviation of 14.80, and for words of low associative potency the mean percentage is 20.97, with a standard deviation of 4.58.

The norms were also used to measure the conventionality of the associations given by the experimental subjects. For each association, the percentage of subjects in the standardizing groups giving the same response was determined. The mean of these percentages provides a measure of the average conventionality of a subject's associations. Separate measures of conventionality were determined for syllables of high and low associative potency, and for words of high and low associative potency. Thus, there were four measures of conventionality for each subject.

To gauge the stability of the measures of individual differences, odd-even reliabilities of the conventionality scores were computed for each group of items. The coefficients of reliability, corrected by means of the Spearman-Brown formula, are presented in Table 1.

Table 1. *Conventionality scores*

Condition of learning	Nonsense syllables						Meaningful words					
	High associative potency			Low associative potency			High associative potency			Low associative potency		
	Mean	S.D.	r_{tt}	Mean	S.D.	r_{tt}	Mean	S.D.	r_{tt}	Mean	S.D.	r_{tt}
Intentional	18.65	7.54	0.64	4.66	3.68	0.57	30.60	10.71	0.94	8.77	2.43	0.82
Incidental	17.08	8.07	0.51	4.24	2.26	0.36	33.01	8.60	0.85	9.12	2.66	0.80

Meaningful words give satisfactorily high coefficients of reliability. The coefficients are considerably lower for nonsense syllables. This difference between the two kinds of materials may be ascribed to the shorter length of the series of syllables—10 items in each subgroup as compared with 50 in the lists of words. When coefficients are estimated for series of syllables comparable in length to the list of words, the differences in reliability all but disappear. The low reliabilities militate, of course, against significant correlations between conventionality scores for syllables and recall. To the extent that such correlations are none the less obtained, they would provide *a fortiori* evidence for the strength of the relationship.

The means and standard deviations of the conventionality scores of the intentional and incidental subjects are presented in Table 1. The results for the two kinds of learners are quite similar and will be considered together. The scores of our experimental subjects conform to the normative classification of the stimulus items but at the same time reveal considerable individual differences among subjects. The conventionality scores for items of high associative potency exceed those for items of low associative potency by a considerable amount. At both levels of associative potency the conventionality scores are higher for meaningful words than for nonsense syllables. An analysis of variance of the conventionality scores is presented in Table 2. (For purposes of this analysis as well as subsequent analyses, the arcsine transformation (Snedecor, 1946, pp. 448–50) was applied to the conventionality scores in order to remove heterogeneity of variance.) While the intentional and incidental groups do not differ significantly, individual differences among subjects are a significant source of variance. Thus, the conventionality scores reliably measure individual differences in verbal dispositions. Both the associative potency of the items and the nature of the material (syllables versus words) are highly significant sources of variance. The significant interaction of these two variables shows that the difference between items of high and low associative potency is reliably smaller for nonsense syllables than meaningful words. The interaction of subjects and materials

significant, i.e. the differences among the subjects are a function of the stimulus materials. There is, on the other hand, no significant interaction of subjects and associative potency. For both sets of materials combined, the differences among subjects are independent of the associative potency of the items.

Table 2. *Analysis of variance of conventionality scores*

Source	D.F.	Mean square	F	P
Between subjects:				
Conditions of learning	1	0.48	0.01	—
Individuals	98	48.39	2.85	<0.01
Within subjects:				
Associative potency	1	21,511.76	1,310.89	<0.01
Materials	1	6,069.43	168.17	<0.01
Conditions × potency	1	2.34	0.14	—
Conditions × materials	1	95.55	2.65	—
Potency × materials	1	376.08	22.17	<0.01
Conditions × potency × materials	1	28.29	1.67	—
Subjects × potency	98	16.41	0.97	—
Subjects × materials	98	36.09	2.13	<0.01
Error	98	16.96	—	—

Table 3. *Mean number of items recalled*

Condition of learning	Nonsense syllables				Meaningful words			
	High associative potency		Low associative potency		High associative potency		Low associative potency	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Intentional	2.86	1.54	2.18	1.48	18.68	4.33	17.94	4.95
Incidental	2.88	1.55	1.90	1.39	17.78	5.62	17.18	5.00

Table 4. *Analyses of variance of recall scores*

Source	D.F.	Nonsense syllables			Meaningful words		
		Mean square	F	P	Mean square	F	P
Between subjects:							
Conditions of learning	1	0.85	0.32	—	34.45	0.85	—
Individuals	98	2.68	1.57	<0.05	40.48	3.87	<0.01
Within subjects:							
Associative potency	1	126.41	73.92	<0.01	22.45	2.15	—
Conditions × potency	1	0.24	0.14	—	0.24	0.02	—
Error	98	1.71	—	—	10.46	—	—

(ii) *The tests of recall.* Table 3 shows the results of the recall tests, and Table 4 presents analyses of variance of the recall scores. The level of recall for words is considerably higher than for nonsense syllables. The over-all mean percentage of words recalled is 35.79, whereas the mean percentage of syllables retained is 24.55. This difference is obtained even though the list of words was five times longer than the list of syllables. In general, the percentage of items recalled is inversely related to the length of the list (Shurrager, 1940), but the difference in the degree of meaningfulness more than overcomes the effects of the length of the series. Since degree of meaningfulness is correlated with associative potency, this finding is consistent with the hypothesis that increase in associative potency favours recall.

Consider next the effects of associative potency within the meaningful and nonsense series. Nonsense syllables of high associative potency are recalled significantly better than syllables of low associative potency. In the case of meaningful words, the difference between the two types of items is in the same direction but it is relatively small and not significant. Thus, intraserial variations in associative potency significantly influence the average recall of nonsense syllables but not of meaningful words. Clearly, associative potency is only one of the determinants of the amount of recall. Syllables of high associative potency are recalled less well than meaningful words of low associative potency. Meaningfulness favours recall, and degree of meaningfulness is correlated with, but not exhaustively defined by, associative potency.

Intentional learners recall more meaningful words and more nonsense syllables than incidental learners. The differences are, however, small and fall short of statistical significance. This result was not unexpected in view of our previous finding (Postman, Adams & Phillips, 1955) that the presence of highly familiar items reduces the difference between intentional and incidental learners. Moreover, the task performed by both intentional and incidental learners—recording a meaningful association to each of the stimulus items—is conducive to learning. Since our main interest is in the comparison of individual differences among intentional and incidental learners it was desirable to equate the two groups as far as possible with respect to the average amount of retention.

Table 5. *Coefficients of correlation between conventionality scores and recall scores*

Condition of learning	Nonsense syllables		Meaningful words	
	High associative potency	Low associative potency	High associative potency	Low associative potency
Intentional	0.14	0.17	0.09	0.23
Incidental	0.11	0.48*	-0.22	0.36*

* Significantly different from zero at the 0.01 level.

(iii) *Analysis of individual differences.* We now turn to an analysis of the relationship between the conventionality of associative responses to the stimulus items and recall. Table 5 shows the coefficients of correlations between the conventionality scores and amount of recall. There are only two coefficients which are significantly different from zero ($P < 0.01$), viz. the correlations for *incidental* learners between conventionality scores on the one hand and recall of words and syllables of *low associative potency* on the other. We have assumed that the conventionality (strength) of explicit differential responses to the stimulus items should have a maximal influence on recall when the items are of low rather than high associative potency and learning is incidental rather than intentional. It appears that a combination of these two conditions is required to produce a significant relationship between conventionality and recall.

Even though the correlations for intentional learners are not significant, it is worth noting that with both types of material the coefficients are somewhat higher for items of low than high associative potency. It is also of interest that with one exception (in intentional learners with items of low associative potency) the positive correlations are higher with nonsense syllables than with meaningful words.

The pattern of correlations supports the conclusion that the relationship between the conventionality of differential responses and recall is a function of (a) the meaningfulness

of the materials, (b) the associative potency of the items, and (c) the set of the learner. The relationship is maximized when the items lack standard meaning, have low associative potency and there is no set to learn. Any deviation from these optimal conditions reduces the magnitude of the relationship.

IV. DISCUSSION

The experimental results support the conclusion that the strength of the differential responses evoked by stimulus items significantly influences recall. Two lines of evidence have been presented: (a) items of high associative potency are recalled better than items of low associative potency, and (b) the conventionality of explicit differential responses given during training is positively correlated with recall. At the same time, our findings point to important limitations upon these generalizations. Associative potency has significant effects on recall with nonsense syllables but with meaningful materials the effects are small and not significant. Degree of meaningfulness is in part a matter of associative potency but clearly these two variables cannot be considered equivalent to each other. Quite apart from the strength of associative responses, meaningfulness has also been related to the number of different associations evoked by a stimulus item (Noble, 1952*a*; Mandler, 1955). This property of verbal stimuli has been shown to influence recall (Noble, 1952*b*), but is not taken into account in our measure of associative potency. We conclude that associative potency is only one of the characteristics of verbal stimuli which determines selective recall.

Similar limitations apply to the relationship between the conventionality of explicit differential responses and recall. Here significant correlations were found only with items of low associative potency under incidental conditions. As we have suggested, the associative context of high-potency items may be aroused implicitly and mediate recall even in the absence of explicit differential responses. The fact that significant correlations are found under incidental but not under intentional conditions makes it clear that motivational factors can maximize the effectiveness of relatively weak verbal habits in mediating recall. This interpretation receives general support from earlier findings that incidental learners show greater selectivity for familiar items than do intentional learners (Postman *et al.* 1955; Postman, Adams & Bohm, 1956).

Our experimental results dovetail with the findings of other experiments which point, directly or indirectly, to the importance of differential responses as mediators of recall. Hovland & Kurtz (1952) have shown that preliminary familiarization with a group of nonsense syllables facilitates subsequent serial learning of these syllables. It is reasonable to assume that at least one of the effects of the preliminary training was to attach differential responses to the syllables which served to differentiate them and to reinstate them in recall. Kurtz & Hovland (1953) have also shown that explicit naming responses at the time of training have significant beneficial effects on the recall of a series of objects. Here the role of differential responses in mediating recall is directly in evidence.

Our results point to certain methodological conclusions: (a) the determinants of selectivity in verbal learning are most readily accessible to experimental investigation under incidental conditions; (b) the role of verbal dispositions in mediating recall can be brought out most clearly in the learning of nonsense series or words of a low associative potency; (c) memory performance can be controlled by preliminary manipulation of the

differential responses evoked by the learning items; (d) individual differences in selective retention can be related to the specific verbal dispositions with which the learner enters the experimental situation.

REFERENCES

- DOLLARD, J. & MILLER, N. E. (1950). *Personality and Psychotherapy*. New York: McGraw-Hill.
- GLAZE, J. A. (1928). The association value of non-sense syllables. *J. Genet. Psychol.* **35**, 255-67.
- HOVLAND, C. I. & KURTZ, K. H. (1952). Experimental studies in rote-learning theory: X. Prelearning syllable familiarization and the length-difficulty relationship. *J. Exp. Psychol.* **44**, 31-40.
- HULL, C. L. (1933). The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.* **45**, 730-4.
- KURTZ, K. H. & HOVLAND, C. I. (1953). The effect of verbalization during observation of stimulus objects upon accuracy of recognition and recall. *J. Exp. Psychol.* **45**, 157-64.
- MCGEOCH, J. A. (1942). *The Psychology of Human Learning*. New York: Longmans Green.
- MANDLER, G. (1954). Response factors in human learning. *Psychol. Rev.* **61**, 235-44.
- MANDLER, G. (1955). Associative frequency and associative prepotency as a measure of response to nonsense syllables. *Amer. J. Psychol.* **68**, 662-5.
- NOBLE, C. E. (1952a). An analysis of meaning. *Psychol. Rev.* **59**, 421-30.
- NOBLE, C. E. (1952b). The role of stimulus meaning (*m*) in serial verbal learning. *J. Exp. Psychol.* **44**, 437-66.
- OSGOOD, C. E. (1953). *Method and Theory in Experimental Psychology*. New York: Oxford University Press.
- PLENDERLEITH, M. & POSTMAN, L. (1956). Discriminative and verbal habits in incidental learning. *Amer. J. Psychol.* **69**, 236-43.
- POSTMAN, L., ADAMS, P. A. & BOHM, A. M. (1956). Studies in incidental learning: V. Recall for order and associative clustering. *J. Exp. Psychol.* **51**, 334-42.
- POSTMAN, L., ADAMS, P. A. & PHILLIPS, L. (1955). Studies in incidental learning: II. The effects of association value on the method of testing. *J. Exp. Psychol.* **49**, 1-10.
- RUSSELL, W. A. & JENKINS, J. J. (1954). *Studies of the Role of Language Behavior*. University of Minnesota, Tech. Rep. no. 11.
- SHURRAGER, H. C. (1940). The measurement of memory on an absolute scale. *Psychol. Monogr.* **5**, no. 234, 21-38.
- SNEDECOR, G. W. (1946). *Statistical Methods*. Ames: Iowa State College Press.
- WOODWORTH, R. S. & SCHLOSBERG, H. (1954). *Experimental Psychology*. New York: Holt.

(Manuscript received 6 September 1956)

RUBELLA DEAF CHILDREN: A RORSCHACH STUDY EMPLOYING MUNROE INSPECTION TECHNIQUE*

By D. MARJORIE BINDON

Department of Psychology, Canterbury University College, N.Z.

With procedure altered to eliminate the deaf subjects' language handicap, the Rorschach test treated by the Munroe Inspection Technique was given to three matched groups of 15-year-old children: (a) thirty-six rubella deaf, (b) fifteen born deaf through causes other than maternal rubella, or who had gone deaf before the age of two, and (c) thirty non-deaf. All quantitative and qualitative results for both deaf groups were strikingly similar: neither cause of deafness, degree of deafness nor the residential factor was found to have had any collateral influence on their performances ($P > 0.1$ in all cases). A criterion score of > 2 statistically selected Munroe checklist entries differentiated both deaf groups from the hearing group ($P < 0.001$ in both cases). It is contended that the rigid personality pattern found to be common to the deaf is not abnormal, but subnormal (immature or undeveloped). It is contended also that their arrested personality maturation is relatively permanent and is not so much due to deafness itself, as to its consequences—a retardation in normal language development which has hindered the process of socialization, and hence full personality development.

I. INTRODUCTION

The present investigation is an attempt to study the personality of deaf children. The Rorschach test was used to reduce to a minimum the language handicap of the deaf subjects; to this end its procedure was also varied. Inter-personal communication was minimized by having no trial blot or initial explanation of the ink-blots, and by holding the short inquiry after each card and determining only the location category for each response. The test was administered individually, but the results were treated by the Munroe Inspection Technique for group administered protocols.

Past Rorschach studies of deaf children have been undertaken by psychologists not qualified to deal with the deaf, and also have serious flaws in experimental design: very small numbers used at each age level, differences noted not tested for statistical significance, no control group, or failure to control R (e.g. McAndrew, 1948*a*, whose average R for his deaf group was 9.0, for hearing group, 30.7), comparison of results with traditional adult norms. Hence the differing findings of previous Rorschach studies (Altable, 1947; Levine, 1948; McAndrew, 1948*a*, *b*) appear in general unreliable, and the present results will therefore not be related to them.

Although rubella deaf children have not been the subject of any published personality research, it is assumed in New Zealand (and overseas, apparently, e.g. Murphy, 1952, p. 119) that they behave differently from other deaf children. From her experience the writer formed the impression that this was not so, but at the same time, that deaf children in general differ from the hearing. Hence, besides the more general aim of attempting to contribute to our knowledge of the personality of the deaf, the present Rorschach study aims to test the following specific hypothesis: that rubella deaf children and

* Based on part of thesis in partial fulfilment of requirements for the degree of Master of Arts in Psychology, University of New Zealand. Thanks are due to Dr T. H. Scott, Canterbury College, for his helpful criticisms of the manuscript.

'non-rubella deaf children' (deaf before the age of two due to causes other than maternal rubella) display a *common* personality pattern, and that they are less mature than their hearing counterparts.

II. PROCEDURE

(i) *Subjects*. (1) Experimental group: thirty-six 15-year-old rubella deaf children. (2) Non-deaf group: thirty hearing children. (3) Non-rubella deaf group: fifteen children who were born deaf through causes other than maternal rubella or who had become deaf before the age of two. The two control groups were paired for sex, educational stream, age, and socio-economic status as gauged by parental occupation, with individuals from the experimental group. An additional criterion of degree of hearing loss was also used to equate the two deaf groups.

None of the deaf subjects had naturally acquired speech and language (hearing loss > 60 decibels), and all were still attending government schools for the deaf in New Zealand where only the oral system of education was in use. The schools are all co-educational and admit both day pupils and boarders.

(ii) *Administration*. The Rorschach test followed a performance intelligence test, and was administered with *S* facing *E* (the writer) to facilitate speech-reading. The following specific instructions were given orally. 'This test I am going to give you is quite different from the first one. There are no right answers and no wrong answers. You can have as long as you like. In this card here [Card I held up, back view facing *S*] many people see all different kinds of things. I want you to tell me all you can see. Do you understand? ... Right. Tell me when you are finished.' [Card handed to *S*.]

The deaf *S*'s were free to supplement their verbal responses with their own sign language, drawing and writing. If *S* gave just one Rorschach response he was told for the first card only: 'Some people can see more than one thing. You saw a —. Can you see anything else? If you cannot that's all right because there are no right answers and no wrong answers.'

This instruction was included in case a deaf *S* might have misunderstood the opening test instructions. The inquiry was held after each card and determined only the location category for each response; it provided also a check against misnaming, or inaccurate recording of the deaf's responses by *E*. After the inquiry, the first card was withdrawn and the second was handed to *S* without comment. If *S* asked whether he could twist the cards he was told he could do as he wished.

(iii) *Scoring*. Though the test was administered individually, it was scored according to the Munroe Inspection Technique instructions for group-administered protocols (Munroe, 1950, pp. 117–34). The total composite checklist scores included both the checklist entries proper and the entries in the checklist left-hand column. The Munroe checklist is based on the Klopfer & Kelley system of scoring, but *I* deviated from their policy of scoring *M* (1946, pp. 109–13) and followed Rapaport (1949, p. 209) because the severe language disability of the deaf could otherwise prejudice their chances of scoring *M* response by making them more prone than the hearing to assign only a name to what they saw in the ink-blot.

III. RESULTS

(i) *Quantitative results*

To avoid capitalizing on possible isolated large chance differences (cf. McNemar, 1949, p. 245) groups were not compared on their individual checklist category scores unless, using χ^2 , their Munroe composite checklist score differences yielded a *P* value of less than 0.01. (χ^2 corrected for continuity, D.F. = 1, cutting point lower median score.)

(a) *Munroe composite checklist scores*

The Munroe composite checklist scores did not significantly differentiate between either the two matched deaf groups, or equated deaf subgroups formed on the basis of degree of deafness, sex and the residential factor respectively (see Table 1). That the Rorschach test did not discriminate between the deaf population supports the writer's view of no behaviour difference existing between rubella and non-rubella deaf children.

On the other hand, in accordance with the writer's hypothesis that deaf children in general differ from non-deaf, the Munroe composite checklist score *did significantly differentiate* the rubella deaf from its matched hearing group ($P < 0.001$, see Table 1).

i) *Comparison of rubella deaf and non-deaf in individual checklist categories*

In view of the above result, the scores were examined for group differences in the separate checklist categories (see Table 2 for summary of findings).

Table 1. *Frequency and χ^2 of Munroe checklist composite scores for matched groups using lower median score as cutting point*

Group	No. of pairs	Frequency	χ^2	P
Matched rubella deaf group	30	27	11.09	<0.001
Non-deaf group		14		
Matched rubella deaf group	15	7	—	—
Non-rubella deaf group		8		
Rubella deaf boys	16	12	0.602	>0.3
Rubella deaf girls		8		
Partially deaf rubella*	18	8	—	—
Total deaf rubella		8		
Residential deaf group	12	6	—	—
Non-residential deaf group		6		

* Arbitrarily defined as those rubella deaf with average hearing loss >60 decibels but <101 decibels in the better ear as tested by a pure tone audiometer in the frequencies 512, 1024 and 2048. Their residual hearing could be utilized to an extent with the help of an individual hearing aid; hence they have acquired language at a faster pace than the totally deaf rubella (term arbitrarily designating the rubella deaf with average hearing loss >100 decibels) and consequently are in separate classes from the totally deaf rubella of similar age and intellectual ability.

Table 2. *Frequency and χ^2 of certain Munroe checklist entries for matched rubella deaf group and non-deaf group*

Checklist entry ...	—				--				---			
Category...	Rub. deaf	Non-deaf	χ^2	P	Rub. deaf	Non-deaf	χ^2	P	Rub. deaf	Non-deaf	χ^2	P
<i>P, Com</i>	21	9	8.06	<0.01	10	2	5.10	<0.05	—	—	—	—
<i>Range</i>	28	21	4.01	<0.05	19	5	11.74	<0.001	16	2	13.41	<0.001
<i>Fk, Fc</i>	19	7	8.21	<0.01	18	3	14.36	<0.001	—	—	—	—
<i>M</i>	27	20	3.54	>0.05	25	13	8.68	<0.01	—	—	—	—
<i>Total Mvt</i>	28	15	11.82	<0.001	25	9	15.27	<0.001	20	4	15.63	<0.001
<i>FC</i>	25	12	10.15	<0.01	20	7	9.70	<0.01	—	—	—	—
	+				++				+++			
<i>Dd %</i>	27	23	1.08	>0.2	17	6	7.05	<0.01	—	—	—	—
<i>F %</i>	30	28			27	11	16.15	<0.001	18	2	16.88	<0.001
	B or B/V											
<i>W</i>	12	0	12.60	<0.001								
<i>O</i>	13	0	14.14	<0.001								
<i>F</i>	14	0	15.75	<0.001								
<i>1 or more B or B/V</i>	17	0	21.01	<0.001								

e) *Generality of certain Munroe checklist entries*

To confirm in the deaf the presence of a distinctive behavioural pattern common to the majority of its population, the following steps were taken. (1) Eight checklist entries, henceforth called 'signs', were selected from Table 2: *Dd* ++, *P, Com* —, *Range* — — —, *F* + + +, *Fc* — —, *Total Mvt* — — —, *FC* — —, and one or more *B* or *B/V* entry: each represents a different checklist category and statistically differentiates the matched rubella deaf group from the non-deaf group ($P < 0.01$). (2) These signs were then applied to individuals in the rubella and hearing group. Compared with four (of thirty) non-deaf, twenty-four (of thirty) rubella deaf children scored >2 signs, this difference being highly significant

($\chi^2 = 24.17$, $P < 0.001$). (3) To check further the differentiating power of these signs, they were applied to the non-rubella deaf group where twelve out of fifteen gained > 2 signs. This criterion significantly differentiated them from the hearing group ($\chi^2 = 17.81$, $P < 0.001$). Thus whereas 80% of the children in both deaf groups obtained > 2 signs, only 13.3% of the hearing did so. Quantitative results therefore suggest that there is a distinctive behavioural pattern common to the deaf children tested (rubella and non-rubella). Furthermore, the individual sign distribution within the two deaf groups was very similar and contrasted markedly with that of the hearing sample (cf. Fig. 1).

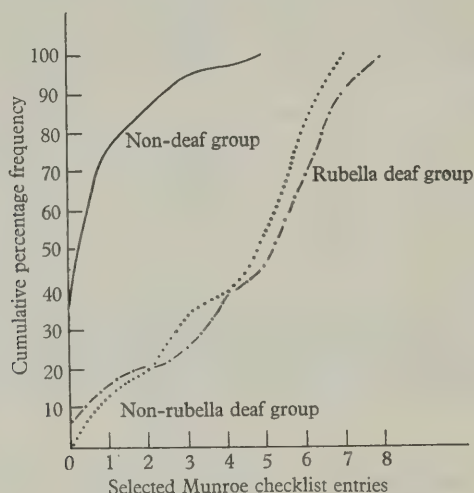


Fig. 1. Cumulative percentage frequency curves for rubella deaf group, non-rubella deaf group and hearing group showing their percentage distribution on eight selected Munroe checklist entries.

(ii) Qualitative findings

This section will discuss only that qualitative information which bears out the described quantitative differences. (The qualitative characteristics of the rubella deaf are those also of the non-rubella deaf.)

(a) Form level quality of rubella deaf and hearing Rorschach responses

It was only in the form level quality of the Rorschach responses that the rubella deaf and non-deaf children differed in direction of deviation from the Inspection Rorschach Record 'norms' of zero; other group differences being in degree of deviation (cf. Table 2). As will be seen later, it is this form level quality of responses that supplies the key, not only to the characteristic Rorschach performances of the rubella-deaf population, but also to those of the hearing children tested. It should be noted first, however, that the checklist data alone (B or B/V entry, seventeen rubella deaf, no non-deaf) do not indicate the full extent of this difference in the form level quality of responses: all but four rubella deaf as against no hearing children, gave $F-$ responses. Further, the deaf protocols usually revealed indiscriminate mingling of $F-$, F and $F+$ responses, and though their productivity was high (av. 30.2, non-deaf av. R 35.2), it was achieved at the cost of a decrease in quality. The finding that none of the non-deaf group produced $F-$ responses points up clearly the outstanding feature of the hearing protocols—their accurate form dominated quality. This 'respect for reality' on the part of the non-deaf revealed itself on the Rorschach records in several ways, for example: cards rejected; W responses replaced by D responses; the Dd category swelled with dr responses (which were not so much responses to very small aspects of the ink-blot, as responses involving the supplementing of D or d responses with parts of the adjacent areas); vague non-committal responses, rated as F , given to those parts of the blot which were themselves indefinite in form. (It is of interest to note that other Rorschach studies of New Zealand children—Malcolm (1951), Spinley (1949), Stroobant (1951)—all found average $F\%$ to be much higher than the $F\%$ norms for American children.)

It has already been mentioned that the deaf population did not have this regard for accuracy. Although, like the non-deaf, they too (1) rejected cards, (2) displaced *W* responses by *D* responses, (3) gave an excess of *Dd* responses (which for the deaf, however, were usually responses to very small blot areas *ld*), and (4) gave vague responses, the explanation of their results differs. It is suggested in the next section that these types of responses are, for the deaf, those which can be given with the minimum of intellectual and perceptual effort and organization.

c) *Certain determinant and content category responses of the rubella deaf*

F—responses. In order of frequency these were mainly (1) *DW* and other confabulatory responses, (2) inaccurate outline responses (due usually to assigning a specific concept to an indefinite form area), and (3) arbitrary responses (generally *Po* responses, or definite form responses drawn in on a formless area; the extreme example here being those few responses drawn in on the white part of the Rorschach card outside and independent of the blot proper).

F responses. Usually evasive form responses such as 'maps' and vague anatomy responses (often small *dr* responses).

F+ responses. The majority were *dd* responses (e.g. foot, mouth, eye, tail). Edge detail responses were next in frequency. Unlike the majority of the hearing group's *de* responses (profile responses), their *e* responses were outlines of such parts of the body as shoulder, waist, mouth and top of head. The many *Do* responses also came into this category.

Content category. Although taken together the rubella deaf responses covered a wide range of associations, their content was predominantly a list of small parts of the body of either animals or humans, anatomy responses, and human object responses (e.g. shoe, belt, pocket, jersey). The hearing group, too, gave many *Hd* and *Ad* responses but they were to larger portions of the blot area and also represented larger portions of the body. Again, the thematic perseveration seen in some of the deaf protocols was rarely present in the non-deaf Rorschach records.

c) *Handling of Rorschach cards*

In the actual handling of the Rorschach cards another important difference between the deaf and hearing subjects was apparent. The non-deaf children handled the cards in a normal manner but the behaviour of many in the deaf group resembled that of individuals of abnormal populations (e.g. purposeless twisting of the cards, 'edging' behaviour which Beck (1946, p. 60) reports occurs exclusively in schizophrenics, tracing outlines of blot areas with their fingers, and, more commonly, 'lapping' the cards which Ames (1952, p. 284) considered to be at least suspect behaviour even for children between two and ten). This last often produced *dr* responses which, though in a sense arbitrary, could not be considered '—' because the blot area selected by this procedure was usually consonant with the configuration of the concept chosen.

IV. DISCUSSION

(i) *Personality pattern of the deaf—abnormal or subnormal?*

The characteristic personality pattern of the deaf with its marked overemphasis on pure form responses and unusual detail responses, very few popular, common, shading and form colour responses, extremely high number of responses confined to the human and animal detail category, poor form level, and a total absence of human, animal, or inanimate movement, resembles those described for such abnormal groups as clinic children (Beck, 1931; Salfeld, 1950; Suarès, 1938), neurotics (Beck, 1937; Rapaport, 1949; Jones, 1948, pp. 118, 119), organic brain cases (Klopfer & Kelley, 1946, ch. xiv) and schizophrenics (Beck, 1937; Rapaport, 1949; Rickers-Ovsiankina, 1938). This finding, however, is not necessarily evidence for the presence of psychopathology in the deaf since their personality configuration also resembles those of young normal children (Ames *et al.*, 1952; Ford, 1946), institutionalized children (Goldfarb, 1944), normal adults under laboratory stress—produced situations (Broida & Thompson, 1954; Cox and Garson, 1954; Eichler, 1951; Siipola & Taylor, 1952), and mental defectives (Beck, 1937; Klopfer & Kelley, 1946, ch. xvi).

This similarity of Rorschach results for such widely differing groups is not due, apparently, to identical thought content, but rather to their similar approach to the handling of the test stimuli brought about, it seems, by a general factor common to them all—the factor of restriction. This factor implies a lack of variability and adaptability and hence implies the presence of rigidity. Where this rigidity is accompanied by a disorganization of functions it is an indicator of abnormality; but where its characteristic feature is a lack of differentiation of functions it is an indicator of subnormality (immaturity or non-adulthood) and is a basic characteristic of undeveloped organisms (see Werner, 1940, 1946).

Since a person's thinking bears the stamp of his personality organization (Mayman, Schafer & Rapaport, 1952), it is the concept formation aspect of the Rorschach test which, together with a knowledge of the past history and present functioning of a specific group, will help to decide whether the rigid behaviour of that group is caused by functional disorganization or lack of differentiation. An analysis of the deaf's conceptual responses shows that, like younger normal children and mental defectives, they have not reached full conceptual development (cf. next subsection). They have been retarded by the restrictive factor of deafness which, in preventing normal language development, has seriously impeded their progressive socialization—a condition which is essential for ensuring full contact with reality and reality testing, and hence for complete cognitive development (cf. Piaget, 1951). The present Rorschach study reveals, therefore, that the restrictive personality configuration of the deaf may be termed a subnormal pattern and indicates that, in comparison with their hearing counterparts, they are functioning on a less mature level of personality integration.

(ii) *Rorschach results in relation to the conceptual processes of the deaf*

The characteristic Rorschach performances of the deaf listed below all required a minimum of intellectual effort and organization and are what the developmentalists would call 'genetically low' responses (Phillips & Framo, 1954): (1) Crude form responses represented fully on the cards (together with an absence of shading responses, where neither texture nor depth is actually on the cards but must be supplied by the subject's imagination (Thomas, 1955), and a total lack of movement responses which are solely a product of the subjects' creative powers and may be termed the genetically highest type of Rorschach response; compare Siipola & Taylor, 1952). (2) Confabulatory, arbitrary and vague responses, or else the most obvious or trivial detail responses (as against only the rare occurrence of responses involving the organization or integration of various aspects of the blot area). (3) Preponderance of personalized or body responses: *Hd, Ad, At, H O*.

To adopt Piaget's terminology, the above conceptual responses show that the deaf thought is still predominantly at the 'transductive stage of development' which 'lies halfway between practical reasoning and truly logical reasoning': the arbitrary responses are examples of 'autistic' thinking, the confabulatory responses of 'intuitive' thinking and the thought content illustrates its 'ego-centric' nature. That is, the Rorschach test reveals that the deaf children's thinking is insufficiently 'decentred' having not yet been transformed by 'socialization' to the highest level of development, the level of 'logical reciprocity'.

(iii) *Rorschach results in relation to the perceptual organizing processes of the deaf*

The above study of the concept formation aspect of the Rorschach test has revealed a retardation in the deaf's conceptual development. But as Hebb (1949) and others have pointed out, the terms 'percept' and 'concept' refer to processes so intimately related that the term 'conceptual development' can cover both and hence, if the first statement is true, the Rorschach test, being a visual test, should also provide evidence of retardation in their perceptual organizing abilities.

In the previous subsection the Rorschach responses characteristic of the deaf were listed, and it was noted that their formulation involved the minimum of intellectual effort and organization. Those same responses illustrate equally well that their operation was on 'the "least" action perceptual tendency' (Hall, 1951, p. 127) and thus are responses to those structural parts of the ink-blot which involved minimum perceptual organization and integration. Again, it may have been the difficulty they experienced in perceptually articulating the Rorschach blots that caused some of the deaf population to exhibit such behaviour as 'blotting', 'edging' and apparently aimless twisting of the cards. Further, their rejection of the Rorschach cards may have been due not so much to their poverty of associations as to their inability to pierce and/or integrate the blot area.

It appears then, in contrast to the findings of Levine (1948) and Murphy (1952), that the deaf, in their visual perceptual abilities at any rate, are not only not superior to their hearing equals, but may actually be inferior to them. The results therefore are against the old hypothesis of sensory compensation, and thus better perceptual organization; rather, they are evidence for 'a theory of correlation which would hold that abilities correlate positively and where one is weak or missing the others are likely to be depressed' (Pintner, 1941, p. 243).

(iv) *Importance of early environmental experiences*

From the age of five the deaf subjects have been exposed to a scholastic programme with its main emphasis on developing their language and speech. Despite this their language development is far behind that of their 15-year-old hearing counterparts (cf. Bakwin, 1950; Drennan, 1941; Ewing, 1941; Pintner, 1941), and it seems that this lag is permanent. Because of the crucial relationship between language development and complete mental development the cognitive development of the deaf, apparently, will also be retarded. Two main factors could possibly account for this: the present teaching methods and techniques used are inadequate to offset completely the consequences of a deprivation of language experiences in pre-school childhood; and/or irrespective of the standard of formal education, the extreme environmental impoverishment of a wordless pre-school childhood has set rather narrow ultimate limits to their future cognitive development. Experimental evidence (cf. Clarke, Heron, Fetherstonhaugh, Forgays & Hebb, 1951; Forgays & Forgays, 1952; Hebb, 1949, ch. 11; Hymovitch, 1952; Thompson & Heron, 1954) supports this last, although it would seem that the quality of formal education will be the main factor in determining the extent to which the deaf can develop within their already restricted limits (cf. Bindon, 1957). Thus if an early restrictive environment in infancy can have such important and relatively permanent effects in

shaping adult behaviour, those concerned with the deaf must direct more attention to the problem of how to present the pre-school deaf child with language experiences at the earliest possible age.

(v) *Wechsler-Bellevue Intelligence Scale Performance Form II and Make-a-Picture Story (MAPS) results for rubella deaf, non-rubella deaf and non-deaf groups*

In addition to the Rorschach test, the Wechsler-Bellevue Intelligence Scale Performance Form II (Wechsler, 1946) and MAPS test (Shneidman, 1948) were administered to all groups.

The results of these tests with the deaf groups, like those of the Rorschach test, revealed that neither cause of deafness, degree of deafness, sex, nor the residential factor had any specific influence on their performances ($P > 0.1$ in the few cases where numerical differences warranted statistical analysis).

On the other hand, the matched rubella deaf group, in comparison with the hearing group, gave on the MAPS test less of Shneidman's 'normal' signs ($\chi^2 = 16.86$, $P < 0.001$) and more of his 'schizophrenic' signs ($\chi^2 = 6.83$, $P < 0.01$); with the results in general revealing their social isolation and conceptual and perceptual incongruity. The intelligence test results did not differentiate between these two groups matched for age, sex, educational stream, and approximate socio-economic status as gauged by parental occupation ($P > 0.1$). The failure to find a statistically significant difference cannot be regarded as necessarily reliable and may be due to the small N , for the mean I.Q. score and all subtest scores were lower for the rubella deaf group. Their consistently lower subtest scores taken with their qualitative performances on this particular test, suggest that they may be less effective in comprehending and sizing up a total situation, in differentiating the essential from the inessential details, and in visual analysis and synthesis. (Wechsler's own subtest interpretations adopted here, 1944, ch. 7.)

The results for these two tests therefore support the Rorschach findings in showing that the all-round cognitive development of the deaf has been retarded.

(vi) *Concluding remarks*

The Rorschach results revealed no behaviour difference between rubella deaf and non-rubella deaf children. They did reveal, however, that the outstanding personality characteristics of the deaf (rubella and non-rubella) were their lack of logical reasoning and their undifferentiated behaviour. This last, it will be remembered, is the basis of that rigidity manifested by immature organisms. Therefore it is suggested that adequate conceptual development may be a necessary pre-condition for the establishment of flexible or differentiated behaviour, which is characteristic of the developed organism. If this is true, any factor that retards cognitive development will also restrict the capacity of the organism to manifest non-rigid and adaptable behaviour. But deafness retards this psychological maturation more than blindness (cf. McAndrew, 1948*a, b*). Hence it is not the actual malfunctioning of one of the major senses but the *consequence* of such malfunctioning that appears to be the more crucial factor in restricting full personality development: deafness is more drastic than blindness because it subjects the deaf to more extreme environmental impoverishment through depriving them of language experiences whose acquisition is the most important factor in the process of adequate socialization—an indispensable condition for complete cognitive development.

REFERENCES

- TABLE, J. P. (1947). The Rorschach diagnostic as applied to deaf mutes. *Rorschach Res. Exch.* **11**, 74-9.
- MES, L. B., LEARNED, J., METRAUX, R. W. & WALKER, R. N. (1952). *Child Rorschach Responses*. New York: Hoeber, Med. Book Dep. Harper.
- AKWIN, R. (1950). The deaf child. *J. Pediat.* **36**, 668-82.
- HECK, S. J. (1931). The Rorschach test in problem children. *Amer. J. Orthopsychiat.* **1**, 501-9.
- HECK, S. J. (1937). Psychological processes in Rorschach findings. *J. Abnorm. Soc. Psychol.* **31**, 482-8.
- HECK, S. J. (1946). *Rorschach's Test. II—A variety of Personality Pictures*. New York: Grune & Stratton.
- BINDON, D. M. (1957). Personality characteristics of rubella deaf children: Implications for teaching of the deaf in general. *Amer. Ann. Deaf.* **102** (in the Press).
- ROIDA, D. C. & THOMPSON, G. C. (1954). The relationship between the Rorschach 'insecurity' hypothesis and children's reactions to psychological stress. *J. Personality*, **23**, 167-81.
- LARKE, R. S., HERON, W., FETHERSTONHAUGH, M. L., FORGAYS, D. G. & HEBB, D. O. (1951). Individual differences in dogs: preliminary report on the effects of early experience. *Canad. J. Psychol.* **5**, 150-6.
- OX, F. N. & SARASON, S. B. (1954). Test anxiety and Rorschach performance. *J. Abnorm. Soc. Psychol.* **49**, 371-7.
- RENNAN, G. L. (1941). The psychology of deafness in children. *Illinois Med. J.* **79**, 227-32.
- ICHLER, R. M. (1951). Experimental stress and alleged Rorschach indices of anxiety. *J. Abnorm. Soc. Psychol.* **46**, 344-55.
- WING, A. W. G. (1941). Defective hearing and mental development (Tests of educational achievement). *Teacher of the Deaf*, **39**, 62-5.
- ORD, M. (1946). *The Application of the Rorschach Test to Young Children*. University of Minnesota Press.
- ORGAYS, D. G. & FORGAYS, J. W. (1952). The nature of the effect of free-environmental experience in the rat. *J. Comp. Physiol. Psychol.* **45**, 322-8.
- OLDFARB, W. (1944). Effects of early institutional care on adolescent personality: Rorschach data. *Amer. J. Orthopsychiat.* **14**, 441-7.
- HALL, K. R. L. (1951). The testing of abstraction, with special reference to impairment in schizophrenia. *Brit. J. Med. Psychol.* **24**, 118-31.
- HEBB, D. O. (1949). *The Organisation of Behavior*. New York: Wiley.
- YMOVITCH, B. (1952). The effects of experimental variations on problem solving in the rat. *J. Comp. Physiol. Psychol.* **45**, 313-21.
- LOFFER, B. & KELLEY, D. (1946). *The Rorschach Technique*. London: Harrap.
- EVINE, E. S. (1948). *An Investigation into the Personality of Normal Deaf Adolescent Girls*. Ph.D. Thesis, N.Y. Univ. Microfilm of complete manuscript. Univ. Microfilms, Ann. Arbor, Mich. Publ. no. 1156.
- MCANDREW, H. (1948a). Rigidity and isolation: a study of the deaf and blind. *J. Abnorm. Soc. Psychol.* **43**, 476-94.
- MCANDREW, H. (1948b). Rigidity in the deaf and blind. *J. Soc. Issues*, **4**, 72-7.
- MCNEMAR, Q. (1949). *Psychological Statistics*. New York: Wiley.
- ALCOLM, O. Y. (1951). *A Comparative Study of Personality Patterns of Maori and Pakeha Children as Revealed by the Rorschach Technique*. Unpublished M.A. Thesis, C.U.C., University of New Zealand.
- LAYMAN, M., SCHAFER, R. & RAPAPORT, D. (1952). Interpretation of the W-B Intelligence Scale in personality appraisal. *An Introduction to Projective Techniques*, ch. 10. Ed. Anderson & Anderson, New York: Prentice-Hall.
- ONS, W. (1948). *Principles and Practice of the Rorschach Personality Test*. London: Faber & Faber.
- UNROE, R. L. (1950). The Inspection technique for the Rorschach protocol. *Projective Psychology*, part II. Ed. Abt & Bellak. New York: Knopf.
- MURPHY, L. J. (1952). *Assessment of Abilities of Deaf Children*. Ph.D. Thesis, Victoria University of Manchester.
- HILLIPS, L. P. & FRAMO, J. L. (1954). Developmental theory applied to normal and psychopathological perception. *J. Personality*, **22**, 464-74.
- LAGET, J. (1951). *Play, Dreams and Imitation in Childhood*. Part III. London: Heinemann.
- UNTNER, R., EISENSEN, J. & STANTON, M. (1941). *The Psychology of the Physically Handicapped*. New York: Crofts.

- RAPAPORT, D. (1949). *Diagnostic Psychological Testing*, vol. II. Chicago: Year Book Publishers.
- RICKERS-OVSIANKINA, M. (1938). The Rorschach test as applied to normal and schizophrenic subjects. *Brit. J. Med. Psychol.* **17**, 227-57.
- SALFIELD, D. J. (1950). An attempt at a numerical evaluation of Rorschach test results. *J. Gen. Psychol.* **43**, 305-11.
- SHNEIDMAN, E. S. (1948). Manual: *Make a Picture Story Test*. Genetic Psychol. Monographs, no. 3.
- SIIPOLA, E. & TAYLOR, V. (1952). Reactions to ink-blots under free and pressure conditions. *J. Personality*, **21**, 22-47.
- SPINLEY, B. M. (1949). *A Study of Two Cultural Groups*. Unpubl. thesis presented for Jacob Joseph Schol., C.U.C., University of New Zealand.
- STROOBANT, R. E. (1951). *A Comparative Investigation of the Personality Configurations of Stammerers and Non-Stammerers*. Unpubl. M.A. thesis, C.U.C., University of New Zealand.
- SUARÈS, N. D. (1938). Personality development in adolescence. *Rorschach Res. Exch.* **3**, 2-12.
- THOMAS, H. F. (1955). The relationship of movement responses on the Rorschach test to the defence mechanism of projection. *J. Abnorm. Soc. Psychol.* **50**, 41-4.
- THOMPSON, W. R. & HERON, W. (1954). The effects of restricting early experience on the problem-solving capacity of dogs. *Canad. J. Psychol.* **8**, 17-31.
- WECHSLER, D. (1944). *The Measurement of Adult Intelligence*. Baltimore: Williams & Williams.
- WECHSLER, D. (1946). *The Wechsler-Bellevue Intelligence Scale. Form II. Manual for Administering and Scoring The Test*. New York: Psychol. Corp.
- WERNER, H. (1940). The concept of rigidity: a critical evaluation. *Psychol. Rev.* **53**, 43-52.
- WERNER, H. (1946). Abnormal and subnormal rigidity. *J. Abnorm. Soc. Psychol.* **41**, 15-24.

(Manuscript received 27 June 1956)

PREVIOUS KNOWLEDGE AND MEMORY

By E. C. POULTON

M.R.C Applied Psychology Research Unit, Cambridge

This paper is concerned: (a) with the relationship between previous knowledge and memory, (b) with differences between recall and recognition, and (c) with set for time of memory test while learning. Subjects listened to 144 short statements, half of which were true and half false. They had to rate each statement on a five-point scale for certainty of their truth or falsehood. A test of memory was then carried out for half of the statements, and another test was given a week later for the other half. In each test, half the statements had to be completed (recall), and each of the remaining statements had to be discriminated from a paired statement which resembled it in a standard way (recognition). While rating the statements, half the subjects did not expect a test of memory at all, a quarter were led to expect an immediate test, and the remaining quarter were led to expect a test in a week. After the first test of memory, no subjects were led to expect a second test. A control group of subjects took a test of 'memory' without hearing the statements.

It was found that a subject was more likely to be certain that a false statement was false, than that a true statement was true. For both true and false statements, memory was best for statements about whose truth the subject was certain, and worst for statements about whose truth he had no idea. The control subject was much less successful than the experimental subject, but tended to score on the same statements. Previous knowledge can thus assist both genuine memory and also guess-work.

With memory for a particular word as a criterion, there was no difference between recall and two-choice recognition when the word was in a statement about whose truth the subject was certain and correct, and memory was tested almost immediately. But recognition was more successful than recall when the word was in a statement about whose truth the subject had no idea, or when memory was tested after a week, although differences were only significant in the latter case.

Recognition was found to be a less sensitive measure of previous knowledge than recall. Success in recognition was determined to a considerable extent by factors intrinsic to the test of recognition. Part of the reduced sensitivity of recognition was attributed to this cause.

Immediate memory was best when a test of memory was expected immediately, and worst when no test was expected, but the differences were not significant. If anything, the trend for memory after a week was in the reverse direction.

I. INTRODUCTION

The primary aim was to determine the size of the effect upon memory, for meaningful statements, of previous knowledge in the field of the statement. The need for this investigation is virtually self-evident for, in a sense, a statement can only have meaning to a subject if he has some relevant previous knowledge; and differences between the learning of meaningful and meaningless materials are well established (McGeoch, 1942). Yet the need for an adequate control of previous knowledge in the design of learning experiments with prose has not always been realized. Effects due to differences in previous knowledge may thus have been ascribed to other causes (Edwards, 1941; Levine & Murphy, 1943; Vernon, 1950; Belson, 1952).

There have been numerous studies relating previous knowledge to success in learning, but most of them have followed a standard pattern: use has been made of nonsense syllables, or single words paired together; and either an artificial amount of previous familiarity has been introduced at the start of the experiment, or familiarity has been

inferred from available group norms, such as those on the association-value of nonsense syllables. In general, prose has been avoided; and when prose has been used, no independent measure has been obtained of each particular subject's everyday experience with each particular item of information contained in the prose.

One subsidiary aim of the present experiment was to investigate two comparisons between tests of recognition and tests of recall. (a) When the same criterion of memory is used for both tests, is there a difference between recall and two-choice recognition? Also, what effect do difficult experimental conditions, such as unfamiliar material and prolonged periods of retention, have upon the size of this difference? No experimental comparison using the *same* criterion of memory for recognition and recall yet appears to have been published. (b) Tests of recognition appear to be less sensitive to changes in the independent variables of the learning situation than do tests of recall (Postman, Adams & Phillips, 1955). Why is this so, and is it possible to improve the sensitivity of tests of recognition?

A second subsidiary aim was to investigate the size of the influence on learning of the set for time of memory test. Previous experiments suggest that memory is best when tested at the expected time (McGeoch, 1942). But the results only appear to be valid when the expected time is immediately after learning. For when, after learning, the subject continues to expect to be tested at some specific time in the future, rehearsal during the interval has clearly not been eliminated as a possible cause of the difference found. The present experiment eliminated this ambiguity.

A group of incidental learners was included for the sake of comparison. Incidental learners tend to do less well on memory tests than intentional learners, but the differences are only marked when the learning task and/or the memory test are difficult (Postman *et al.* 1955).

II. METHOD

Four lists, each of thirty-six statements, were recorded on magnetic tape. A statement was read every 10 sec. One of the most familiar statements was: 'The stern may be the rear of a steamer'. The least familiar was: '*Sinanthropus* was dug up in Australia'. One of the statements which obtained the most qualified ('probable') ratings was: 'A tablespoonful is four teaspoonfuls'. About 30% of the statements could be classified as scientific, 20% as historical, and 15% as geographical. They were selected to be sufficiently different from one another in meaning to minimize the number of confusions between them.

The statements ranged in length from 3 to 9 words, with an average of 6.1 words. Each started with a title, e.g. the stern, and contained two key words, e.g. rear and steamer. About 20% of the titles contained a word which was probably unfamiliar to half or more of the subjects (see Results). No two key words were the same. About 80% of them were included in the commonest 2000 words of the English language (Thorndike & Lorge, 1944). (Slight modifications were made in cases where English and American usage differ, e.g. petrol and gasolene.) The remaining 20% were included in the commonest 7000 words. Only one such less common word was included in any one of the statements. These statements were divided equally between the four lists, and were distributed randomly within lists.

About 8% of the statements were phrased in the negative. They were distributed randomly between and within lists. Half the statements in each list would generally be considered to be true, the other half false. About 95% of the true statements were simple statements of fact, whereas only about 40% of the false statements were either simple statements of fact containing a negative, or in some sense the opposite of the truth. The remaining false statements were not related to the truth in any simple uniform way. The order of true and false statements within lists was randomized.

(i) *Self ratings of previous knowledge.* The experiment was carried out in three parts. The first part was presented as a new kind of general knowledge test. The statements were played over loudspeakers. The subject had a list of the titles of the statements in front of him. After each title there were five boxes. For half the subjects they were labelled in order: certainly true, probably true, no idea, probably false, certainly false.

certainly false. For the other half of the subjects the boxes were labelled in the reverse order. After hearing each statement, the subject had to put a tick in the relevant box. The instructions were phrased so as to prevent him from being either over-cautious or over-rash. The difficulty of the statements was so graded that about one-third of them should have fallen into each of the three categories: certain, probable, and no idea, and the subject was told this. (Difficulty was estimated in a pilot experiment with twenty-four other subjects.) If the subject did not hear a statement, he was instructed to leave a blank, not to place a tick in the box labelled no idea.

The experimenter started by reading out six examples. In each case, he stated how he would respond, and why. This was followed by a practice of twelve statements, played over the loudspeakers. The subject's responses were subsequently checked and discussed by the subject and experimenter together. If there was to be a warning that a test of memory would follow the test of general knowledge (see below), it was then given. There was a break of about one minute after the first two lists of statements had been heard.

(ii) *Immediate test of memory.* Immediately after the test of general knowledge, there was a test of memory for the first two of the four lists of statements. To twenty-four subjects, the incidental learners, this came as a complete surprise. Twelve subjects had been warned to expect an immediate test of memory, and the remaining twelve had been warned to expect a test in a week's time, but not immediately. In no case was the nature of the test of memory disclosed in advance.

The titles of one list of statements were presented on an otherwise blank sheet, and the subject had to complete as many of the statements as he could. He was told to give the original wording if possible, and at least to give the correct meaning. The statements of the other list were then presented in full. Each statement had a similar statement typed above or below it, from which it had to be distinguished. Half the originals were typed first, and half second, the order of pairs being randomized within lists. The subject was made to guess if he was not sure. But he was told that he should be able to decide in almost all cases without guessing. Half the subjects performed the recall test first, half the recognition test. Unlimited time was available for both tests.

In the test of recognition, the alternative version of a statement had the same title as the original. Half the alternative versions had one different key word, and the other half had both key words different. Each of these subgroups was divided into three, according to meaning. The meaning of the alternative was either more or less the same as the original, in some sense the opposite, or as far as possible unrelated. The alternative statements were carefully paired with the originals. Thus a statement which contained a less-common key word was paired with a similar statement. In addition, the alternative statements were selected to make alternative lists, which were similar to the original lists in all the respects required by the experimental design. Thus the alternative lists could also have been used for the test of general knowledge, and this was originally the intention.

(iii) *Delayed test of memory.* This was carried out one week later, and was similar to the immediate test. It used the last two of the four lists of statements heard. This test came as a complete surprise to all subjects, and thus minimized rehearsal. The subjects who had been warned to expect a test in a week's time, had been told at the time of the immediate test that the immediate test was actually to be their test.

(iv) *Order of lists.* Twelve of the twenty-four incidental learners heard the four lists of statements in the order 3 4 1 2. The remaining thirty-six subjects all heard the lists in the order 1 2 3 4. The lists were constructed to be of approximately equal difficulty. Of each subgroup of twelve subjects, six had to recall lists 1 and 3 and recognize lists 2 and 4. The other six had to recognize lists 1 and 3 and recall lists 2 and 4. Memory for list 1 was always tested before memory for list 2, and memory for list 3 before list 4. This design confounds the difference between immediate and delayed memory with the order of hearing the lists.)

(v) *Control group.* A group of twenty-four subjects had to take recall and recognition tests without having heard the statements. In a preliminary training period each subject copied down the six examples, and the nature of the statements was explained to him. Afterwards he had to complete ('recall') the twelve practice statements. His attempts were examined by the experimenter and himself, and the actual statements used in the practice were then read to him. In the test which followed he had first to complete the statements of one list, and then to select ('recognize') one of each pair of statements of one other list. He was told to judge between the pairs of statements on their merits, and not just to guess at random. Conditions were otherwise the same as for the memory tests of the experimental groups. Twelve of the subjects worked with lists 1 and 2, the remainder with lists 3 and 4. Of each subgroup, half completed one list and the other half completed the other list.

(vi) *Subjects.* The seventy-two subjects were all naval ratings, aged between 17 and 27 years, with a median age of 20. Their scores on intelligence test AH 4 (Heim, 1955) ranged from 43 to 116 with a median of 72. Most had left school at 15, and had been apprentices or had taken miscellaneous jobs before joining the Navy. They were tested in groups of six. The experimental groups were of approximately similar composition.

(vii) *Scoring.* Recall was normally scored by two methods (see Table 2 or Table 3). (a) In one method a statement was considered to be correctly recalled if the two key words were given more or less correctly. The remaining words, and the meaning of the recalled statement, were neglected in this method. The exact form of the key words (e.g. adjective, noun, or verb, singular or plural, present or past tense) was also neglected, since this depended on the structure of the sentence. Only about 60% as many statements would have been scored as correctly recalled by this method if the exact wording of the original had been required. (b) In the second method, a statement was considered to be correctly recalled if the meaning was more or less correct. The recalled version had to reflect the meaning of one of the two key words, and not to alter too grossly the general sense of the other key word. Excess information was not penalized either if it did not alter too grossly the meaning of the original statement, or if the meaning of the original was ridiculous anyway. In order to satisfy the experimental design, it was necessary to ensure that the criteria used for the various statements were equally strict. However, a certain degree of apparent arbitrariness cannot be avoided in matching differences in meaning. If the recalled version had had to reflect the meaning of both key words, only about 80% as many statements would have been scored as correct by this method.

In comparing recall with recognition, the criterion used was remembering one particular key word. This was the key word which was changed in the set of alternative versions with more or less the same meaning as the original statement, and with one identical key word. In the test of recall, the criterion for the key word was the same as that described in (a) above. (The key word had only to be recalled sufficiently correctly for the experimenter to be reasonably certain which one of the two alternative words used in the test of recognition was intended.) This criterion of memory restricted the comparison to one-sixth of the statements.

(It is not possible to compare recall and recognition on memory for meaning alone, without altering in the test of recognition the wording of both the original statement and the alternative version, for an alternative version with a different meaning necessitates some change in wording. And a change in wording as well as in meaning favours recognition, unless the 'original' statement has also undergone the corresponding change in wording—see Discussion.)

(viii) *Calculations.* An index of general knowledge was computed by subtracting the certain and probable true-false ratings which would generally be considered to be incorrect, from those that would generally be considered to be correct. A recognition score was also computed, by doubling the number of errors in the test of recognition, and subtracting this product from the greatest possible score. (The recognition score is the closest approximation which can be obtained to the amount of genuine memory shown in the test of recognition. The underlying assumptions are (a) that errors are wrong guesses, and (b) that corresponding to every wrong guess there was on average a guess which was right, since there were only two possible choices and the subject was forced to select one of them. On this view, doubling the errors eliminates on average the correct guesses.)

The significance of the difference of each pair of means was assessed separately (a) against the variability between subjects, and (b) against the variability between statements. The level of the less significant of the two tests was taken. This test was almost always (b). Two-tailed tests were used except where stated. As compared with more conventional methods, this method tends to underestimate significance. For Tables 2 and 5 the difference between paired proportions was assessed, since the incidence of the scores of previous knowledge varied as well as the memory score. A simple sign test was used for each pair of proportions, and the numbers of positive and negative outcomes were compared by χ^2 .

The means of the experimental subjects were computed from the results of all forty-eight subjects. However, the levels of significance were generally computed only from the results of the thirty-six subjects who heard the lists in the order 1 2 3 4. Exceptions to this rule are the significance levels in Table 1, which are for all forty-eight subjects, and others specifically mentioned in the text. (The memory scores of the twelve subjects who heard the lists in the order 3 4 1 2 could not be combined with the remainder when carrying out test (b).)

III. RESULTS

(i) *Previous knowledge.* The distributions of the self ratings of previous knowledge are shown in Table 1. The subject was more likely to be certain that a statement which would generally be considered to be false was false, than that a true statement was true. The order of the five boxes (see Method) made no difference to the ratings. About 20% of the statements received 'no idea' ratings from half or more of the subjects. Practically all these statements contained an unfamiliar word in the title. They accounted for about 55% of all the 'no idea' ratings. The extreme values of the index of general knowledge for different subjects were 20 and 68%. For brevity, the data on individual differences will not normally be given, but the scatter was always large.

Table 1. *Truth of statement and self rating of previous knowledge*

	Percentage of statements rated				
	Certain		Probable		No idea
	Correct	Incorrect	Correct	Incorrect	
True statements	31	5	27	12	25
False statements	39	3	20	11	27
Difference	-8	2	7*	1	-2

* This difference is significant at the 0.001 level of confidence. It is significantly different from the difference for certain correct at the 0.01 level.

(ii) *Memory.* Table 2 gives the memory scores for the different self ratings. The statements about which the subject was certain were remembered best, and the statements about which he had no idea were remembered least well. This trend was shown by both the true and the false statements. In addition, the statements which received the conventional true-false rating ('Correct' in Table 2) were remembered better than the statements which received the corresponding unconventional rating ('Incorrect').

Table 2. *Self rating of previous knowledge and memory*

Criterion of memory	Self rating of previous knowledge				
	Certain		Probable		No idea (%)
	Correct (%)	Incorrect (%)	Correct (%)	Incorrect (%)	
Memory after 30 min.					
Two key words reproduced*	56	41	44	33	16
Meaning more or less reproduced†	89	76	76	65	33
Recognition score‡	92	84	90	94	79
Memory after 1 week					
Two key words reproduced§	18	16	13	11	2.2
Meaning more or less reproduced¶	45	30	31	26	9
Recognition score**	76	62	70	66	40

* Combined correct and incorrect certain-probable, $P=0.05$. Probable correct different from probable incorrect at 0.05 level, and from no idea at 0.01 level.

† Probable correct different from both certain correct and no idea at 0.01 level.

‡ Probable incorrect-no idea, $P=0.01$.

§ Certain correct-no idea, $P=0.01$.

¶ Certain correct-probable correct, $P=0.01$.

** Certain correct-no idea, $P=0.05$.

The meanings of the true statements were rather better recalled after 30 min. than the meanings of the false statements ($P=0.02$). There was no difference between memory for the different kinds of false statements. Only about 5% of the recalled versions of the false statements had a meaning in some sense opposite to that of the original. The corresponding figure for the true statements was 3%. Mishearings in the original test of previous knowledge appeared to account for less than 1% of the errors in recall.

Memory after 30 min., for the 20% of statements which contained an unfamiliar word in the title, was about the same as the results in Table 2 for the statements rated 'no idea'. After a week memory for the statements with an unfamiliar word in the title was rather worse in every case than the results in Table 2 for the 'no idea' statements.

In the test of recall after 30 min. none of the forty-eight subjects obtained full marks by either criterion. Only 3% of the statements to be recalled by eighteen subjects always had the two key words reproduced correctly, and only 18% of the statements always had the meaning more or less correctly reproduced. In contrast, in the test of recognition after 30 min., 12% of the subjects, and 65% of the statements to be recognized by eighteen subjects, produced no errors. After a week 17% of the statements were still always recognized correctly, although none of the subjects obtained full marks for recognition.

(iii) *Set for test of memory.* By all three criteria of memory, immediate memory was best when the immediate test of memory was expected, and worst when no test was expected. Expecting a test in a week gave intermediate results. If anything, the trend for memory after a week was in the reverse direction, although the differences were smaller. However, none of the differences due to set for test of memory was significant with the stringent tests of significance used ($P>0.05$).

(iv) *Guess-work.* Table 3 shows the memory scores after a week for all statements combined, together with the scores for the guesses of the twenty-four control subjects who did not hear the statements. Also given are the τ correlations by statements between the scores of the control subjects, and the scores after a week of the twenty-four incidental learners. The memory scores after a week were significantly better than guess-work although all subjects tended to score on the same statements in the tests of statement completion.

Table 3. *Memory after 1 week compared with guess-work*

Criterion of memory	Memory after 1 week (%)	Guess-work (%)	τ by statements, memory after 1 week (incidental learners)- guess-work
Two key words reproduced	12	1*†	0.30‡
Meaning more or less reproduced	29	9*§	0.37¶
Recognition score	64	4.5*	0.09

* Memory after 1 week (incidental learners only)-guess-work, $P=0.001$.

† Guess-work different from zero at the 0.05 level.

‡ $P=0.02$.

§ Guess-work different from zero at the 0.001 level.

¶ $P=0.001$.

For the control subjects only 6% of the completions of the false statements had a meaning in some sense opposite to that of the original. When the two alternatives in the test of statement selection had the same meaning, the control subjects selected the correct alternative 25% more often than would be predicted by chance ($P=0.001$).

(v) *Recognition.* The effect on recognition score (see Calculations) of the relationship between the original and the alternative statement is shown in Table 4. Over 40% of the errors in recognition after 30 min. were attributable to 8% of the statements. These statements were typed below the alternative statement, contained one of the same two key words, and had more or less the same meaning as the alternative.

Table 4. *Effect of recognition test on recognition score*

Relationship between original and alternative statement	Recognition score after			
	30 min. (%)	<i>P</i>	1 week (%)	<i>P</i>
Original typed: above	96	0.05	70	0.05
below	81		54	
Identical key words: 1	82	0.05	52	0.001
	93		74	
Meaning: same	79	0.05	80	>0.1
	90		78	
	94		68	
opposite				
different				

In both the immediate and the delayed tests of recognition, the statements paired with the different kinds of alternative all showed the same trend with previous knowledge. The nature of the alternative only influenced the mean recognition score, and the size of the trend. In the test of recognition after a week, the statements paired to alternatives with only one key word changed were the most sensitive statistically to differences in previous knowledge. For these statements the difference between certain correct and no idea was significant at the 0.02 level (cf. Table 2).

Table 5 compares two-choice recognition with recall, when the same criterion of memory was used in both tests. Recognition was either as successful as recall, or more successful.

Table 5. *Recall compared with recognition*

Criterion of memory	Self rating of previous knowledge			
	Certain correct* (%)	Probable correct (%)	No idea* (%)	All ratings combined† (%)
Memory after 30 min.				
One specific key word reproduced	70	44	22	50
Recognition score when alternative statement had only this key word changed	69	78	34	64
Memory after 1 week				
One specific key word reproduced	33	19‡	3.6	22
Recognition score when alternative statement had only this key word changed	72	72	21§	55

* After 1 week recall-recognition, $P=0.05$, with a one-tailed test.

† After 1 week recall-recognition, $P=0.01$.

‡ Certain correct-probable correct, $P=0.05$, with a one-tailed test.

§ Certain correct-no idea, $P=0.05$, with a one-tailed test.

(vi) *Experimental material.* There was no difference in any respect between the statements containing the less common key words and the remaining statements. The four lists of statements gave similar mean ratings for familiarity. In addition, the memory scores of the two groups of incidental learners showed that the two halves of the material were about equally difficult to remember.

(vii) *Incidental learning and intelligence.* The τ correlation of the total score on intelligence test AH4 (Heim, 1955) with the index of general knowledge was 0.55 for the subjects who expected a test of memory, and 0.36 for the subjects who did not expect one ($P=0.01$ or better). In order to determine the direct correlation of intelligence with memory, the correlations of the index of general knowledge with intelligence on the one hand and with memory on the other had to be partialled out. When this was done for the incidental learners, the total on AH4 correlated 0.30 with immediate memory, and 0.37 with delayed memory. The corresponding values for the subjects who expected a test of memory were 0.01 and 0.04. Recall of the two key words was used as the measure of memory. Thus with subjects who expected a test of memory, intelligence was reflected principally in the index of general knowledge. Whereas with the incidental learners intelligence was about equally reflected in the index of general knowledge and in memory. These results are in line with the view that intelligence is closely related to incidental learning, and thus to general knowledge, but is not so closely related to the ability to learn by intention (Carr & Kingsbury, 1938, p. 373).

IV. DISCUSSION

(i) *Previous knowledge assisted memory.* Table 2 shows that, when the subject was certain and correct about the truth of a statement, he was more likely to recall its meaning than when he was less certain of its truth ('Probable Correct' in Table 2). This difference was significant at the 0.01 level both when memory was tested after about half an hour and when it was tested after a week. Similar trends were found when only the presence or absence of the two simple key words in each reproduction was considered, and in the test of recognition, although these trends were not significant. The differences are interpreted as follows: Degree of certainty about the truth of a statement was related to the amount of previous knowledge in the field of the statement, and amount of previous knowledge was related to success in the tests of memory.

Table 2 also shows that statements about whose truth the subject was certain and correct, were recalled after 30 min. 2.7 or 3.5 times better than statements about whose truth he had no idea, depending on the criterion used. After a week the ratios were 8.2 and 5.0. On this extreme comparison both criteria of recall showed significant differences at both time intervals, and recognition showed a significant difference after a week. The memory scores for the 20% of statements which contained an unfamiliar word in the title were much the same as the scores in Table 2 for the statements rated 'no idea'. In fact, about 55% of all the 'no idea' ratings went to these statements. Hence the difference in memory between the statements rated 'certain' and 'no idea' is to some extent analogous to the difference found between an old 'stimulus' and a new 'stimulus' in experiments on the learning of nonsense syllables by the method of paired associates.

It was also found that statements which would generally be considered to be true were rather better recalled (though not necessarily better recognized) than false statements although the subject was more likely to be certain in his rating of false statements (Table 1). Most knowledge is probably stored in a positive rather than a negative form. Thus the false statements were presumably less directly related to previous knowledge than were the true statements. (The difference was not due simply to recalling true

meanings which were the opposite of the correct false meanings. For the proportion of opposite meanings recalled was about the same for the true and false statements.)

(ii) *Previous knowledge assisted statement completion by guess-work.* Table 3 shows that the control subjects, who had to do the tests of 'memory' by guess-work, scored significantly better than zero on statement completion ('recall'), although not on statement selection ('recognition'). In addition, they tended to score on the same statements as the experimental subjects. This was to have been expected, since the guessing of the control subjects in completing the statements was presumably directed by previous knowledge.

Previous knowledge may also have directed the guessing of the experimental subjects, as well as assisting them in learning and recall. Table 3 suggests that guess-work could not have been much help in spotting the actual wording of the statements to be recalled. This was because many statements did not contain the customary key words, owing to the requirement that the same key word must not appear in two statements. But guess-work may have helped considerably with the meanings of the statements, especially in the test after a week. In the present experiment the information in different statements was more or less independent. In normal prose guessing may be a more important factor in successful 'memory'. For in prose there is generally no limitation on the incidence of key words, and consecutive meanings are not independent of each other.

A small but significant 'recognition' score was obtained by guess-work when the two alternative statements had the same meaning. The correct alternatives tended to have the more common combinations of words. This small bias in the experimental design should not have occurred.

(iii) *Relevance to some previous memory experiments.* Edwards (1941) read a controversial passage to two groups of students with widely different political views. In a subsequent test of recognition, he found that both groups tended to remember the points which corresponded to their political outlook better than the points which conflicted with it. Following an hypothesis derived from Sherif (1936), he concluded that political frames of reference acted selectively on the learning of, and memory for, the controversial material. But in his design differences in previous knowledge between the two groups were not balanced. There was an additional finding that the over-all memory score of the group containing the more politically experienced students was higher than the score of the group containing the less politically experienced students. This strongly suggests that previous knowledge did influence the results. Differences in previous knowledge congruent with the political frames of reference cannot therefore be excluded as the principal determinant of the differences found. A similar criticism applies to the experiment of Levine & Murphy (1943).

The present results also suggest that an uncontrolled test of memory is not an appropriate measure of the comprehensibility of a passage. This measure has recently been used on educational talks for adults broadcast by the B.B.C. (Vernon, 1950; Belson, 1952). But this measure does not distinguish between what was known before hearing the passage, and new information which has been acquired by listening. Comprehension is confounded with previous knowledge. Some confounding is inevitable, for memory is, as it were, loaded *against* new knowledge (Table 2). But the use of a control group of subjects similar to the experimental group, which takes the same test without having listened to the passage (cf. Table 3), will at least indicate what could have been achieved by previous knowledge alone.

(iv) *Truth and certainty.* Table 1 shows that the subject was more likely to be certain in his rating of a statement which would generally be considered to be false, than in his rating of a true statement. His greater caution with the true statements was probably not due to having available less relevant previous knowledge. For the proportions of incorrectly rated statements, and the proportions of statements about which he had no idea, were practically the same for the two types. The true statements must therefore have seemed more ambiguous to the subject than the false statements. (About half both types of statement appeared to the experimenter to be ambiguous in some degree. This was because they were selected to be short and simple, and were thus oversimplifications. Brevity and simplicity without ambiguity is the target of all designers of so-called objective tests (Ruch, 1929; Vernon, 1940).)

(v) *Variables which affected the relationship between success in recognition and success in recall.* In comparing the relative difficulty of tests of recall and recognition, the same criterion of memory must be used for the two tests. The criterion used in Table 5 was remembering one particular key word (see Scoring). The justification for using recognition score to represent the amount of genuine memory shown in the test of recognition has been outlined above (see Calculations). With statements about whose truth the subject was certain and correct, recall and two-choice recognition were about equally successful in the test after 30 min. But as conditions became more difficult, with less familiar statements and/or delayed memory, two-choice recognition was the more successful.

No previous experiment comparing recall with recognition appears to have used the same criterion of memory for the two tests. Differences in criterion inevitably favour recognition. For the easiest test of recall, the reproduction of a single item more or less correctly, corresponds to the most difficult test of recognition, the selection of the correct version when the alternative(s) have only this single item different. The more that is demanded concurrently in a test of recall, the *more difficult* it becomes. Whereas the more that is changed concurrently in a test of recognition, the *easier* it becomes (see Table 4).

As an illustration, we can consider an experiment by Postman, Jenkins & Postman (1948), who claimed to have demonstrated that recognition was easier than recall. For recall they demanded perfect reproduction of three-letter nonsense syllables, and for recognition they presented four choices like the following: QUN, MUN, JIC and JIV. But as compared with recall, recognition was too easy, for two letters (UN, JI) were varied concurrently. The criterion of memory which corresponds to the perfect recall of a three letters, demands a recognition test with at least the following eight choices: ABC, AB3, A2C, 1BC, A23, 1B3, 12C and 123.

(A test of recognition can vary along two distinct dimensions. First, it can test memory for more than one item. And secondly, for each item to be tested, there can be more than one alternative. If memory for each item is to be tested independently (as is necessary when making the correct comparison between recognition and recall), and every alternative is to vary independently of every other alternative, the total number of alternatives will be $(r+1)^n$, where r is the number of alternatives for each item, and n is the number of items for which memory is to be tested. Thus in the modified Postman, Jenkins & Postman experiment, with one alternative for each of three items, the total would be $2^3=8$. Recognition becomes more difficult with more alternatives, whether the number is increased to test memory for each item independently, or simply to increase the total number of choices for each item (Postman, 1950).)

(vi) *Why recognition was a less sensitive measure than recall.* In Table 2 recognition after 30 min. showed only one rather atypical significant difference with previous knowledge, whereas both the measures of recall showed the predicted significant differences. For short-term memory, recognition was thus a less sensitive measure of previous knowledge than was recall. A somewhat similar finding has been reported by Postman *et al.* (1955) for nonsense syllables. One reason for the relatively low sensitivity of recognition may have been that the test was too easy after 30 min., for 12% of subjects and 65% of statements escaped without any errors. Table 2 shows that after a week recognition did just give the over-all significant difference ($P=0.05$) with previous knowledge which was predicted. This difference became more significant ($P=0.02$) when only the more difficult pairs of statements containing one identical key word were considered.

Table 4 suggests a second reason for the relatively low sensitivity of recognition, which applies to both short-term and longer-term memory. Success in recognition depended to a considerable extent upon two factors which were specific to the test of recognition, in addition to the experimental variables. These factors were the order of presentation of the old and new items, and the similarity of the new items to the old. If a test of recognition is to be made as sensitive as possible, either the influence of these factors must be held constant, or the part of the variance resulting from them must be isolated. (In the present experiment the number of statements in each subgroup was too small to give any significant differences with previous knowledge after 30 min., and the form of the results ruled out analysis of variance.)

Table 3 indicates a contributory cause of the difference in sensitivity between recognition, and recall of the meaning of the statements. Guessing may have somewhat increased the 'recall' score of the meanings of the more familiar statements, especially in the test after a week. However, Table 3 suggests that guessing did not appreciably affect the recall score of the two key words in each statement.

The influence of Sir Frederic Bartlett, F.R.S., is shown clearly in this paper. Prof. T. A. Ryan pointed out the need for an adequately designed experiment to investigate the influence of set for time of memory test upon learning. Prof. P. E. Vernon loaned me his copy of his 1950 report. Mr M. Stone gave advice on statistical methods. The subjects were supplied by the Royal Navy. Financial support from the Medical Research Council is also gratefully acknowledged.

REFERENCES

- BELSON, W. A. (1952). An inquiry into the comprehensibility of 'Topic for Tonight'. B.B.C. Listener Research Report, no. 1080.
- CARR, H. A. & KINGSBURY, F. A. (1938). The concept of ability. *Psychol. Rev.* **45**, 354-76.
- EDWARDS, A. L. (1941). Political frames of reference as a factor influencing recognition. *J. Abnorm. Soc. Psychol.* **36**, 34-50.
- HEIM, A. W. (1955). *Manual for the Group Test of General Intelligence AH4*. London: National Foundation for Educational Research.
- LEVINE, J. M. & MURPHY, G. (1943). The learning and forgetting of controversial material. *J. Abnorm. Soc. Psychol.* **38**, 507-17.
- McGEOCH, J. A. (1942). *The Psychology of Human Learning*. New York: Longmans, Green.
- POSTMAN, L. (1950). Choice behavior and the process of recognition. *Amer. J. Psychol.* **63**, 576-83.
- POSTMAN, L., ADAMS, P. A. & PHILLIPS, L. W. (1955). Studies in incidental learning. II. The effects of association value and of the method of testing. *J. Exp. Psychol.* **49**, 1-10.

- POSTMAN, L., JENKINS, W. O. & POSTMAN, D. L. (1948). An experimental comparison of recall and recognition. *Amer. J. Psychol.* **61**, 511-19.
- RUCH, G. M. (1929). *The Objective or New-Type Examination*. Chicago: Scott, Foresman.
- SHERIF, M. (1936). *The Psychology of Social Norms*. New York: Harper.
- THORNDIKE, E. L. & LORGE, I. (1944). *The Teacher's Word Book of 30,000 Words*. New York: Teachers College, Columbia University.
- VERNON, P. E. (1940). *The Measurement of Abilities*. London: University of London Press.
- VERNON, P. E. (1950). The intelligibility of broadcast talks. B.B.C. Listener Research Report, no. 2328.

(*Manuscript received 26 May 1956*)

MEASURES OF SUBJECTIVE PROBABILITY

ESTIMATES OF SUCCESS IN PERFORMANCE IN RELATION TO SIZE OF TASK

By JOHN COHEN, E. J. DEARNALEY AND C. E. M. HANSEL

Department of Psychology, University of Manchester

The question, which arose in an earlier investigation, of the transposability of estimates made on different ranges ('out of five', 'out of ten', etc.) is subjected to experiment. The degree of transposability is found to be a function of age.

I. INTRODUCTION

By subjective probability we understand a measure of mental states of uncertainty which extend from complete uncertainty at one extreme to complete certainty at the other. These states are implicit in all human behaviour, in every perception, thought, judgement, feeling and action, and they assume particular significance in all manner of risk-taking, including gambling, in judgements about the outcome of future events, and in our orientation to the future as partly expressed in our expectations, hopes and fears. One of the aims of the study of subjective probability is to make these implicit states of mind explicit, so that they can be measured.

In our recent studies of risk-taking (1956*b*) we have measured subjective probability (ψ) by asking each person to guess how many times he would succeed in a task out of five attempts. The number five was used because each subject was actually to make five attempts, so that his estimates could be compared with his performance. In this particular situation it would have been tedious for the subjects to make more than five attempts at each level of difficulty. In other situations, subjects have been asked how many times out of ten or out of one hundred they thought they would succeed. Measures of subjective probability were obtained by taking the ratio of the number of expected successes to attempts.

This procedure raises the question whether different values for subjective probability would be obtained if the subject had to express his successes in relation to a larger or smaller number of attempts. An experiment was therefore designed to answer this question. Children were asked to estimate the number of times they would succeed in throwing a ball to hit a target when the number of attempts was set at each of the following values: 1, 5, 10, 100, or 1000. This enabled us to see whether the ratio of estimated successes to attempts would be constant regardless of the number of attempts. These particular values were chosen because they were within the range of the experience of the school-children to be studied. For example, the range 0 to 10 is associated with football, and 0 to 500 with the game of cricket. Children in English schools are generally well informed on matters relating to these two national sports. The targets were varied in size so as to yield eight levels of difficulty and ensure a diversity of values of subjective probability on each range.

II. PROCEDURE

Four classes of school-children consisting of 62 children aged 9-10 years, and 68 children aged 13-14 years acted as subjects. They had to imagine they were throwing a table-tennis ball to hit a target through circular aperture 4 ft. away. The aperture and target were placed on an easel in front of a class and one child was asked to make a trial throw in order to illustrate the hypothetical task. This trial throw was made when the aperture was 2 ft. 3 in. in diameter. The attempts always succeeded. A series of apertures was presented with diameters of eight different magnitudes, viz. 2.25, 1.75, 1.50, 1.25, 1.00, 0.75, 0.50, and 0.25 ft. Each child received a booklet of six pages. On each page the child recorded his estimates of his expected successes in aiming at each of the apertures in a given number of attempts (1, 5, 10, 100 or 1000), a different range being set on each page. The questions, printed at the top of the page, were of the following form: 'You have a thousand shots at the target. How many times do you think you will hit it?' Each class of children was divided into five subgroups, each subgroup receiving the tasks in a different order. The order in which the five ranges were presented was different for each of the five subgroups as set out in Table 1.

Table 1. *Order of presentation of ranges*

Subgroups of children				
I	II	III	IV	V
0-1	0-5	0-10	0-100	0-1000
0-1000	0-100	0-5	0-10	0-1
0-100	0-10	0-1	0-1000	0-5
0-5	0-1000	0-100	0-1	0-10
0-10	0-1	0-1000	0-5	0-100

The apertures were presented in random order with respect to size. Two of these apertures (0.75 and 1.50 ft.) were presented twice to provide a measure of the reliability of the estimates. Each child thus made fifty estimates.

At the foot of each page, after ten estimates had been made, the subjects answered the following question: 'How many of the ten estimates you have made do you think would be correct if you *actually* carried out the task each time and not just *imagined* that you did it?' Finally, on the sixth page, the children answered the following questions given and explained verbally: (i) Which of the five kinds of question did you find easiest to answer?; (ii) Which did you find the hardest to answer?; (iii) Which of the five questions (or ranges) would you use if you were explaining to a friend how hard it is to throw the ball at this target through the aperture in front of you? (The aperture was set at a diameter of one foot.)

The five different ranges were written on the blackboard to make it easier for the children to reply to the above questions

III. RESULTS

The median estimates of success made by the two age-groups are shown, expressed as ψ -values, in Figs. 1 and 2. Medians rather than means are shown since, in some cases, children made obvious errors because they sometimes carried over one range to the next page. Rather than reject their estimates entirely, and thus introduce a factor of selection, we have employed the median which reduces the effects of large isolated errors on the results. The first point that emerges from Figs. 1 and 2 is that subjective probabilities, as measured by estimates of success, vary systematically with the number of hypothetical attempts. The effect is similar in both age-groups but smaller in the older group. Estimates of success decline as the number of hypothetical attempts increase.

The second point is that subjective probabilities of success vary with the difficulty of the task, which is presumably related to the size of the target. In the younger group estimates of success decline progressively in all ranges as the task becomes harder. It may

that the range of 'out of a thousand' is beyond the competence or experience of some of these 9-year olds. In the older group the estimates also decline, but not to the same extent; they seem more stable and less affected by the range on which they are expressed. The extent to which the ranges 'out of five' and 'out of a hundred' are equivalent and therefore transposable is brought out in Table 2. There is a marked tendency in the older group to move towards the 'expected' values.

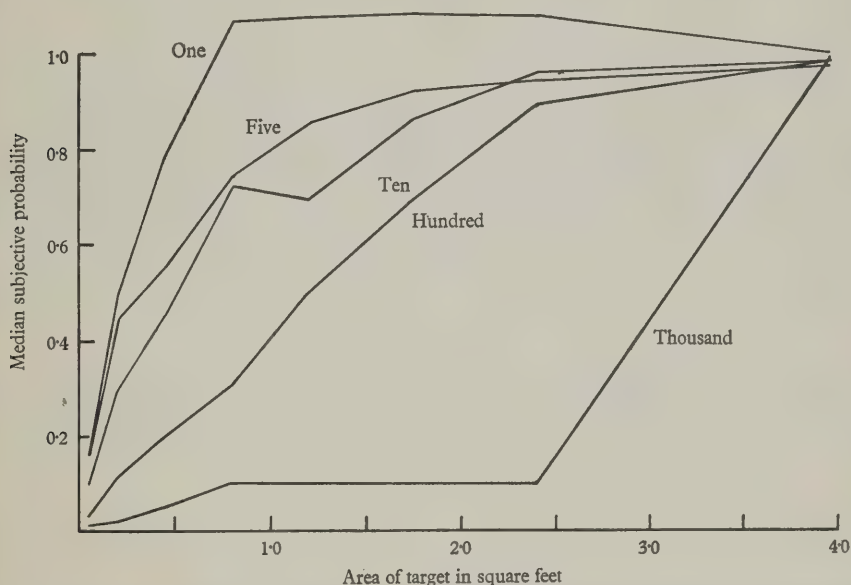


Fig. 1. Subjective probability of success in relation to area of target for different scales of judgement. ($n=62$ children age 9+).

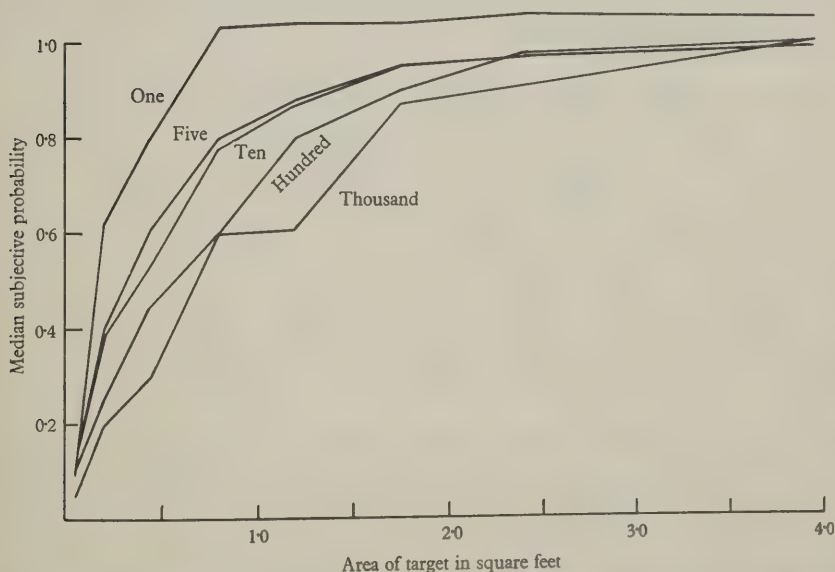


Fig. 2. Subjective probability of success in relation to area of target for different scales of judgement ($n=68$ children age 13+).

Table 2. Median estimates on the 'hundred' scale corresponding to estimates on the 'five' scale for all targets

Age	'Out of five'					
	0	1	2	3	4	5
9-10	2	8	20	30	35	90
13-14	8	15	30	45	75	99
'Expected' values	5	20	40	60	80	95

We are now in a position to give an affirmative answer to the question posed in the introduction above, as far as our subjects are concerned and under the conditions of the experiment. The fact that the range effect declines with age justifies the supposition that its effects would be slight in corresponding estimates made by adults. A further point of interest arose in an analysis of the distribution of estimates on the hundred scale. They show marked preferences for particular values varying with the age of the children. The values favoured by the younger group, after allowing for the differences in frequencies due to the difficulty of the task were 0, 1, 2, 3, 4, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 99 and 100. (A value is called 'favoured' if its corrected frequency exceeds 1% of the total frequency.) The older children favoured the values 1, 5, 10, 15, 20, 25, 30, 35, 40, 50, 60, 70, 75, 80, 85, 90, 95, 99 and 100. These represented 70% of the values used by the younger children and 80% of the values by the older ones. The preferences suggest that in both age-groups the intervals of discrimination tend to be finer towards the ends of the range and coarser in the middle. This appears to indicate that ψ_2 values are high when they relate to either extreme of the ψ_1 scale and low when they refer to the middle region of the ψ_1 scale, as we found in experiments on intensity of belief (1956*a*).

The consistency of the individual estimates may be judged from a comparison of the medians of the two sets of values obtained from the target diameters (0.75 and 1.50 ft) presented twice. These are shown in Table 3. The twenty pairs of medians (in Table

Table 3. Median values of estimates of success at targets presented twice

Range	Size of diameter							
	0.75 ft.				1.50 ft			
	Age		Age		Age		Age	
	9-10	13-14	9-10	13-14	9-10	13-14	9-10	13-14
Out of:								
One	0.78	0.78	0.90	0.65	1.07	1.10	1.04	1.05
Five	3.3	2.4	3.8	2.9	4.6	4.7	4.8	4.8
Ten	4.5	4.7	5.9	5.2	8.5	8.8	9.4	9.6
Hundred	20	17	45	43	68	70	95	90
Thousand	101	31	328	300	106	155	894	800

are fairly close to each other. We may safely assume that estimates of success at other target diameters are equally reliable. We have also calculated the reliability coefficients between the sixteen pairs of estimates (omitting the 'out of one' range). Eleven of the sixteen coefficients exceed 0.7. In both age groups the majority of children chose 'out of one' as easiest, 'out of a thousand' as most difficult, and 'out of ten' as the rarest they would use to explain the difficulty of the task to a friend. The ψ_2 values, which sign

an individual's degree of confidence in his estimates, range from 6.5 to 8.1. They differ significantly for the various ranges and show a slight tendency to decline as the range becomes larger. The older group give consistently larger values at each range.

IV. CONCLUSIONS

Our main conclusions may be briefly summarized as follows. (i) Subjective probabilities of success at a given task, as measured, tend to decline as the number of hypothetical attempts at the task become larger.* This effect is much less marked in the 13-14 age-group, than in the 9-10 age-group, and we may infer that it would be negligible in adult estimates. (ii) On all ranges, the children show preferences for certain values. These indicate a finer discrimination towards the ends of the range than in the middle region; the older children show a finer discrimination than the younger ones. A range 'out of a hundred' is warranted for the age groups 13-14, although they, like the younger children expressed a preference for the 'out of ten' range.

REFERENCES

- COHEN, J. & HANSEL, C. E. M. (1956*a*). *Risk and Gambling*. London: Longmans Green.
COHEN, J. DEARNALEY, E. J. & HANSEL, C. E. M. (1956*b*). Risk and hazard. *Operat. Res. Quart.* **10**, 67-82.

(*Manuscript received 5 June 1956*)

* A similar effect is found in subjective probabilities of failure. The present paper should be read in conjunction with a subsequent study by the writers on the relationships between subjective probabilities of success and failure. (The Addition of Subjective Probabilities, *Acta Psychol.* 1956, **12**, 371-80.)

THE QUALITATIVE INVESTIGATION OF INTELLIGENCE

BY ANN SHEDDEN MCCOLSKEY

University of Florida

I. INTRODUCTION

A battery of seven intellectual tasks, designed for the qualitative examination of intelligence, was administered to thirty-five (experimental) children of superior ability and thirty-five (control) children of average ability, matched for age and other variables, and selected by a set of multiple criteria, including teachers' opinions, grades, and intelligence and achievement tests.

Under the conditions of this study, superior intelligence was characterized, with significant frequency, by a variety of qualities, including: coherence and productivity of informal thought processes; a facility for abstracting conceptual principles as against 'situation-bound' responses; conceptual flexibility; perseverance and independence of structure (or aid) in complex problems; attentiveness and spontaneous perception of incongruities. It is felt that this approach to intelligence can ultimately produce a more fertile, inductive quantification, in dimensional terms.

The undivided attention afforded the quantification of intellectual behaviour by psychologists since the early part of this century—reinforced by the lay enshrinement of the I.Q.—has unhappily stunted the growth of the logically complementary, analytic, or qualitative approach. Practical exigencies fostered neglect of the controlled investigation of the more basic, complex, intellectual processes (which should *precede* quantification of an adequately complex and representative sort) and favoured the leap to immediate quantification of intelligence, on a more or less intuitive basis (cf. Wechsler, 1944, p. 13; Hearnshaw, 1955, p. 316), and in terms of the more superficial and palpable, intrinsically metrical aspects of intellectual functioning—speed, accuracy, retention, extent of vocabulary, and the like.

The inevitable result of intelligence testing of this sort is an array of paradoxes. The clinician's primary skill, for example, in reporting intelligence test performances, lies in his ability to resurrect, *intuitively*, the complex qualitative, or processual, context for the occasional intrinsically quantifiable response. Again, the expedient, *a priori* equation of intelligence with (palpable, intrinsically metrical) verbal skills (cf. Terman & Merrill, 1937, p. 5; Anastasi & Foley, 1954, *passim*) has given birth to a paradoxical environmental dependency-relationship which finds intelligence fluctuating according to cultural, subcultural, racial, regional, and rural-urban differences; deafness; foster home situations; and a host of other conditions, and fetters research by its inability to transcend such inequalities in opportunity for verbal development.

If one refuses the ultimate tautology of the traditional approach ('Intelligence is what the tests measure'), the alternative is a reorientation towards *representative* quantification, *inductively* derived from analytic observations of the complex mental processes which accompany the occasional, directly quantifiable response, and which fundamentally differentiate individual performances. The initial problem, which this study attacks, is one of devising a standardized framework for the observation and recording of basic behaviour patterns in a variety of intellectual tasks; of categorizing the results, and determining significant differences in preference for the various categories—the qualitatively different modes—of intellectual operation between differing levels of intellectual ability.

II. PROCEDURE

A qualitative battery. The battery devised for the present study (described in full elsewhere, McColskey, 1955) includes seven tasks, of varying length and complexity. Through each task is threaded a standardized administrative procedure for every conceivable level of response; this facilitates categorization, yet permits complete freedom of response. Each task also controls or by-passes the more usual, superficial measures of quantitative success, in order to minimize the effects of differential training and experience, as well as of motivational differences based on a consciousness of success or failure,* and to isolate the more 'native', covert, qualitative differences in behaviour between two levels of mental ability. No subject 'passes' or 'fails' a task in this battery.

(1) In a free-association 'Word-naming' task, a three-minute flow of words is recorded verbatim, under set toward continuity, not speed (as in the Binet task), and is examined to determine whether significant differences in logical coherence complement the traditionally remarked differences in productivity, or the gross number of words.

(2) A tripartite 'Vigotsky' task employs the 22 blocks of the Vigotsky test, but departs considerably from the traditional administrative form. The basic Vigotsky problem requires the subject to sort the blocks, which comprise varying combinations of colour, shape, height, and surface-area, into four coherent groups. The difficult solution, which involves the interactive concepts of height and size, is realized through a trial-and-error mental process which is concretely reflected in the block manipulations. The traditional measures of time and of errors (Hanfmann & Kasanin, 1937) are lamentably sterile, as they are neither demonstrably reliable (Semeonoff & Laird, 1952, p. 99), nor representative of the stage-by-stage conceptual processes reflected by the test, as Heidebreder (1948) and others lament. The writer has constructed a systematic observational schema for the test, which, among other things, provides for presentation of successive series of four blocks of a shape, in a predetermined order. (Previously, clues have been offered in a hit-or-miss fashion, and an error of administrative judgment 'may alter the entire course of the subject's subsequent performance' (Semeonoff & Laird, 1952, p. 99).) The problem is thus progressively structured for each subject, until all achieve 'success'.

A similar sequence of progressive aid facilitates categorization of the verbal explanations of the correct groupings, and contrasts the initial, spontaneous, offerings with the levels reached with prompting. Categorization of explanations was aided by suggestions from Hanfmann & Kasanin (1937); Ray (1936); Semeonoff & Laird (1952). Two preliminary Vigotsky tasks also have been introduced, which first require the subject to sort the blocks on the simplest possible conceptual basis (by colour or by shape), and then to re-sort by an alternative concept (shape or colour). These transitional tasks provide a uniform frame of reference for the Vigotsky problem, and capture qualitative differences in *simple* conceptual operations, and in conceptual *flexibility*, to contrast with differences in the *complex* conceptual processes of the Vigotsky problem proper.

(3) Four short 'Absurdities' (adapted from Terman & Merrill, 1937) are presented in the guise of memory items, with discreet questioning to test for qualitative differences in the spontaneous appreciation of incongruities.

(4) Three simple 'arithmetic' problems, termed 'Confusions', are insoluble because of inadequate information, but are designed to mislead the unwary into offering a specious 'answer', and to search out qualitative differences in mental alertness or attentiveness.

(5) Three Rorschach cards are utilized to test for qualitative differences in modes of perceptual organization—in 'global', or unitary, 'whole' responses against differentiated 'whole' percepts—which would resemble certain age-changes reported in several Rorschach developmental studies (Hertz & Libert, 1944; Meyer & Thompson, 1952; Stavrianos, 1942).

(6) Qualitative differences in 'Definitions' of nine, very simple, common words are examined, *apart from* the more usually investigated, culturally influenced, differences in *range* of vocabulary. (Categorization of 'Definitions' was facilitated by Feifel and Lorge's (1950) categorization of ontogenetic changes in quality of definitions.)

(7) Six pairs of simple, common words, selected from the Binet and Wechsler scales, comprise a 'Similarities and Differences' task, which aims to reveal qualitative differences in the complementary capacities for perceiving relationships and for differentiation.

The subjects. The battery was administered individually to 70 fourth- and fifth-grade children of two north Florida towns, in the summer of 1954. Thirty-five children were of superior intelligence; the other

* An after-inquiry of likes and dislikes revealed *unanimous* and enthusiastic endorsement of this 'new kind of test'.

35 constituted a control group, of average intellectual ability. In each group, there were 18 boys and 17 girls; 13 fifth-graders and 22 fourth-graders, and the socio-economic status of the average group approximated that of the superior group. Teachers' opinions were employed as preliminary criteria for selection; recent intelligence test scores, school marks, and achievement test scores over a 2 to 3-year period were then considered, *interdependently*. (In general, none of these crude ability indicators was given independent weight, but each child passed cutting points on all four criteria, or on three of four, evidence on the fourth were borderline, and other data were extraordinarily impressive.)

The composition of each group is described fully elsewhere (McColskey, 1955), but may be sketched here briefly. On the California Test of Mental Maturity, which routinely had been administered in schools, for example, superior I.Q.'s ranged from 125 to 147, with a mean of 134; comparable statistics for the average children were 90-110, and 102. Overlap was minimized but not prohibited, for scores along the remaining two objective criteria, in view of the relatively restricted range, and varying school and teacher standards; borderline cases here were assessed in terms of the consistency of all other data, as stated above. Thus, in honour-point equivalents, the superior group ranged from 3.0 to 4.0, around a mean of 3.7; corresponding data for the control group were 1.5 to 3.6, and 2.5. In achievement test scores, superior children ranged from 0.5 years below to 1.9 years above actual grade level, and were accelerated 1.0 year, on the average, while similar figures for the control group were 1.3 (below) to 0.6 (above), and 0.3 years' retardation.

III. RESULTS

Treatment of results consisted of categorization of varying behaviour-reactions and response-patterns within each task, and comparison of experimental and control group frequency distributions by the χ^2 technique. (Marginal distributions provided the 'expected' proportions; neighbouring levels were combined as necessary, to ensure five or more 'expected' subjects; Yates' correction for continuity was employed where appropriate. It should be noted that the χ^2 treatment in the present paper represents a refinement of that employed in the original study.) Within any selected dimension of behaviour, categories are presumed to be exhaustive; are built on concrete referents, not subjective judgments; and are inductively derived from (not imposed on) the obtained protocols. Complete data are presented in the original study; a more succinct summary of findings is offered here.

Word-naming. A brief analysis of the logical coherence within protocols revealed significant differences ($P < 0.02$) between experimental and control groups, in favour of the former. Coherence was defined in the simplest and most objective fashion possible, for this preliminary study, by classifying protocols according to the number of clearly thought sequences (three or more obviously related words, comprising a generic sequence as: 'red, white, blue', or an associative train, as: 'lion, roar, zoo') as opposed to isolated or questionably related words. Group distributions may be illustrated by the fact that 40% of the superior group and only 11% of the average employed six or more sequences, while, at the other extreme, 31% of the average protocols and 9% of the superior were almost entirely disjunctive in quality, exhibiting two or fewer sequences. The expected group differences were found in the more commonly explored, complementary dimension of productivity (total number of words), with one-third of the superior group again 9% of the average naming 60 or more words, and 26% of the average against 0.0% of the superior limited to 39 or fewer words ($P < 0.01$). Further investigation (of mean coherence per productivity level, and vice versa) indicated that these twin aspects of thought processes in flux are less than perfectly correlated, especially among average children, that a high productivity level does not necessarily imply notable coherence of thought.

Vigotsky, preliminary tasks. There were no group differences in the initial selection on a simple principle—colour or shape—for sorting the Vigotsky blocks, in (preliminary

Task I: approximately two-thirds of both groups first chose shape as the preferred concept. Interesting and significant group differences in the initial reactions to the instructions for this relatively novel and primitive conceptual task were discovered, as indicated in Table 1. (Initial reactions were not systematically observed for all subjects on subsequent Vigotsky tasks. Statistics computed for those which were observed indicated increasing differentiation of groups with increasing task complexity, and a progressive downward shift in the locus of greatest differentiation.) Nearly one-third of the superior subjects impatiently began Task I before the examiner completed her instructions, while only one average subject did so. On the other hand, 43 % of the average subjects demonstrated some difficulty in comprehension or acceptance of the task, in contrast to 14 % of the experimental group.

Table 1. *Initial reactions to Vigotsky, preliminary tasks*

Initial reaction	Frequencies		
	Observed*		Expected
	Av.	Sup.	
A. Impatience: too-early start	1	11	6.0
B. Immediate start	19	19	19.0
(B). Immediate start, but questions during instructions	5	2	10.0
C. Delayed start (sits in thought, or asks several preliminary questions)	7	2	
D. No start; requires repetition of instructions	3	1	
Totals	35	35	35.0

* $\chi^2 = 13.334$; $P < 0.01$ (D.F. = 2: $P < 0.01$, $\chi^2 = 9.210$; $P < 0.001$, $\chi^2 = 13.82$).

Table 2. *Final (prompted) explanations of Vigotsky sortings, preliminary tasks*

Final explanation	Frequencies		
	Observed*		Expected
	Av.	Sup.	
A. Concept abstracted and named	25	33	29.0
B. Groups separately described, correctly	10	1	6.0
C. Irrelevant, vague, partial descriptions of groups	0	0	
D. Fails to explain	0	1	
Totals	35	35	35.0

* $\chi^2 = 4.298$ (corrected for continuity); $P < 0.05$ (D.F. = 1: $P < 0.05$, $\chi^2 = 3.841$; $P < 0.02$, $\chi^2 = 5.41$).

Classification of subjects according to the mode of grouping the blocks revealed insignificant differences in Task I, although 26 % more superior than average subjects preferred an uncomplicated sorting into (5) colours or (6) shapes, while 26 % more average than superior children produced a differentiated (one or more of the five or six primary groups subdivided in some way); an apparently confused; or no sorting.

Similarly, classification of the initial (unprompted) explanations of the (correct) sorting revealed no significant group differences. A slight initial superior proclivity (54 % against 37 %) for efficient abstraction and naming of the conceptual principle involved, however, did become significantly pronounced in the final explanations, with the admixture of systematic prompting by the examiner. As shown in Table 2, virtually all of the superior children were able to reach this level with prompting, while more than one-fourth of the control group remained 'situation-bound'—limited to description of the specific block-groups.

Upon instructions to shift to an alternative elementary sorting principle, in (preliminary) Task II, significant group differences in response again were demonstrated, as seen in Table 3. Nearly three-fourths of the superior youngsters, in contrast to approximately one-third of the average group, evinced independent conceptual flexibility, at this primitive level.

Table 3. *Shifting simple concepts, Vigotsky, preliminary tasks*

Flexibility	Frequencies		
	Observed*		Expected
	Av.	Sup.	
A. Spontaneous, appropriate re-sorting	12	25	21.0
B. Re-sorts with aid, after initial:			
B ₁ , selection of new, but inappropriate concept	2	3	6.5
B ₂ , perseveration with original concept	6	2	
B ₃ , vague, unclear sortings	5	0	
B ₄ , aimless or no activity	4	2	7.5
C. Rigidity: any of B ₁ -B ₄ , but no re-sorting, in spite of aid	6	3	
Totals	35	35	35.0

* $\chi^2 = 12.562$, $P < 0.01$ (2 D.F.: $P < 0.01$, $\chi^2 = 9.210$; $P < 0.001$, $\chi^2 = 13.82$).

Vigotsky, Task III. The Vigotsky problem proper provided some of the most intriguing and clear-cut group differences in the entire battery. Categorization of subjects by the level of structure required to achieve the correct Vigotsky sorting very significantly ($P < 0.001$) differentiated the two groups. With no more than two levels of structure, for example, 74% of the superior children reached the solution, in contrast to 29% of the average subjects. Conversely, even with maximal structure (17 of the 22 blocks correctly placed for the subject), 37% of the average group against 3% of the superior group still could not perceive the solution.

Group differences in the time taken to reach solution reaffirmed the marked superior tendency toward perseverance in an unstructured problem situation, and average inclinations toward inertia, with fairly constant demands for structure. Superior subjects took significantly *longer* than average ($P < 0.05$) to reach solution, with 83% of the latter solving the problem in fifteen minutes or less, in contrast to 49% of the superior group (These data cast some light upon the anomalous results of studies which have assumed speed of solution to constitute a positive measure of success in the test.)

Once all subjects had achieved (or been shown) successful solution of the Vigotsky problem, explanations of the correct sorting—both the initial, spontaneous offerings, and the final, prompted, verbalizations—still served to differentiate the two groups to a remarkable degree, as Table 4 demonstrates. Three-fourths of the superior children, in contrast to but 14% of the average, were able spontaneously to perceive and verbalize the twin concepts, in some form. Nearly one-half of the average subjects, against 11% of the superior, could not offer even a partially adequate explanation, without assistance. Yet the introduction of prompting exaggerated, rather than minimized, group differences in response, as 89% of the superior children achieved succinct abstraction of the twin concepts, or at least accurate group descriptions, with aid, while 77% of the average group still could not advance beyond successive, 'situation-bound', comparison of pairs of block-groups.

Table 4. *Initial and final explanations, Vigotsky*

Explanation	Initial (spontaneous) explanation (frequencies)			Final (prompted) explanation (frequencies)		
	Observed*		Expected	Observed†		Expected
	Av.	Sup.		Av.	Sup.	
A. Double concept abstracted and named ('It's by size and height')	1	3	15.5	2	24	13.0
B. Accurate description of each group, concepts used interactively ('These are big and tall; these are small and tall')	1	17		6	7	6.5
(B.) Successive comparison of groups, concepts used alternately ('These are taller than these, but bigger around than those')	3	6		5	0	7.0
C. Single concept abstracted and named ('It's by size')	11	5	10.0	5	4	
D. Description of each group, with variations on a single concept ('Big, medium, small, tiny')	4	0		9	0	8.5
E. Irrelevant, vague description of groups ('They're different shapes...')	7	3	9.5	3	0	
F. None offered	8	1		5	0	
Totals	35	35	35.0	35	35	35.0

* $\chi^2=25.593$, $P<0.001$ (2 D.F.: $P<0.001$, $\chi^2=13.82$).† $\chi^2=38.264$, $P<0.001$ (3 D.F.: $P<0.001$, $\chi^2=16.27$).Table 5. *Absurdities*

Recognition of incongruity	Item 1. Frequencies			Item 2. Frequencies			Item 3. Frequencies			Item 4. Frequencies		
	Observed*			Observed†			Observed‡			Observed§		
	Av.	Sup.		Av.	Sup.		Av.	Sup.		Av.	Sup.	
Immediately precise	4	15	12.5	3	15	10.0	6	21	16.5	15	27	21.0
Circumlocutory but, actually correct	2	4		0	2		2	4		7	4	5.5
Possible vague notion, never explicitly stated	4	3	22.5	0	0	25.0	0	0	18.5	6	0	8.5
One	25	13		32	18		27	10		7	4	
Totals	35	35	35.0	35	35	35.0	35	35	35.0	35	35	35.0

* $\chi^2=10.516$, $P<0.01$ (1 D.F.: $P<0.01$, $\chi^2=6.635$; $P<0.001$, $\chi^2=10.83$).† $\chi^2=13.720$, $P<0.001$ (1 D.F.: $P<0.001$, $\chi^2=10.83$).‡ $\chi^2=16.568$, $P<0.001$ (1 D.F.: $P<0.001$, $\chi^2=10.83$).§ $\chi^2=9.010$, $P<0.02$ (2 D.F.: $P<0.02$, $\chi^2=7.82$; $P<0.01$, $\chi^2=9.210$).

After all subjects had offered or had been provided A-level explanations (at the conclusion of progressive, level-by-level, demonstration and discussion), instructions to demonstrate a firm grasp of the Vigotsky principle by re-sorting the blocks into the correct block-groups without aid still significantly differentiated experimental and control subjects ($P<0.05$), in terms of errors made. Sixty-nine percent of the superior children, for example, against 31% of the average, re-grouped without error.

Absurdities. Spontaneous appreciation of the incongruities involved in the four 'memory' items was consistently and significantly characteristic of the superior group, rather than the average, as Table 5 demonstrates. On all four items, superior children strongly predominated (by 32% or more) at a level of immediate sensitivity to the

incongruity, while, on three of the four items, the large majority (71 % or more) of average children betrayed no recognition whatsoever. Summarizing these results, a re-categorization of subjects according to the number of ‘Absurdities’ explicitly recognized, immediately or eventually, very significantly differentiated the two groups ($P < 0.001$), with 54 % of the experimental group and but 9 % of the control responsive to at least three of the four items.

Confusions. As can be seen in Table 6, reactions to two of the three *Confusions* significantly distinguished experimental and control groups, with the vast majority of the former (69 % or more) explicitly or implicitly alert to the subtle omissions in these ‘arithmetic’ problems, while the bulk of the latter (69 % or more) successfully were led astray. The first and most devious problem apparently deceived or confused both groups, although trends here were consistent with the pattern elsewhere. An overall, summary classification of subjects by the number of ‘Confusions’ resolved, explicitly or implicitly, significantly differentiated the two groups, with 60 % of superior children alert to at least two of the discrepancies, and 51 % of average children perceiving none.

Table 6. *Confusions*

	Item 1. Frequencies			Item 2. Frequencies			Item 3. Frequencies		
	Observed*		Expected	Observed†		Expected	Observed‡		Expected
	Av.	Sup.		Av.	Sup.		Av.	Sup.	
Attentiveness									
A. Refusal of solution, because of inadequate information	2	7	5.5	7	19	13.0	5	22	
B. Solution, pending further information	2	0		3	5		1	2	
(B). Solution, supplying missing information	7	11	9.0	1	4	6.5	0	0	
C. Deep thought; vague perception of discrepancy; no solution attempted	5	7	6.0	0	0		2	0	
D. Immediate false solution	15	10	14.5	24	7	15.5	27	11	
E. Blank; total confusion; immediate refusal of problem (‘I’m no good at arithmetic’)	4	0		0	0		0	0	
Totals	35	35	35.0	35	35	35.0	35	35	

* $\chi^2 = 4.834$, N.S.
† $\chi^2 = 16.784$, $P < 0.001$ (2 D.F.: $P < 0.001$, $\chi^2 = 13.82$).
‡ $\chi^2 = 18.900$, $P < 0.001$ (1 D.F.: $P < 0.001$, $\chi^2 = 10.83$).

Perceptual organization. Qualitative differences in the mode of organization of ‘whole’ Rorschach responses were significant on Card 1 ($P < 0.01$), but not on Cards 2 and 7 although the direction of differences was the same in all three cases. (The latter two cards produced very few ‘global’ responses, in either group.) On Card 1, approximately three-fourths of the average subjects, against 36 % of the superior, reported a ‘global’ ‘whole’ percept only, while 64 % of the superior children, against 26 % of the average responded either with a differentiated ‘whole’ or with both differentiated and ‘global’ ‘whole’ percepts.

Definitions. Of the five ‘Definitions’ (‘bicycle’, ‘puddle’, ‘nonsense’, ‘join’, and ‘compare’), in which the factor of range of vocabulary was excluded as desired (i.e. in which all subjects were thoroughly familiar with the words), in order to isolate more or less ‘native’, qualitative definitional differences, four significantly differentiated experimental and control groups ($P < 0.001$ in three cases; < 0.01 in one), and all five items demonstrated the same pattern of differences. Categorization was, of course, specific to

the 'Definition', but a typical overall progression emerged, which is illustrated in Table 7. In all five cases, superior subjects most often (from 44 % to 91 %) offered either concise synonyms, or fuller descriptive explanations, while average definitions most typically (from 45 % to 70 %) were either overly vague and general or overly narrow and particular, or else wholly amorphous.

Similarities and differences. No formal results in this task can be presented because the fact that several subjects offered more than one response (χ^2 demands independence of frequencies in each class) precluded computation of probability figures, for group differences. In general, group differences in response were marked—somewhat more so in 'Similarities' than in 'Differences'—and resembled those found in 'Definitions'. The extraordinary heterogeneity of the item-categorizations in this task, however, defied extraction of a clear-cut overall progression, although superior responses tended to comprise precise generalizations of basic similarities or differences, while average responses tended towards indefiniteness, irrelevance, or narrow particularization. In response to the first 'Similarity', 'car-boat', for example, 83 % of the superior and 42 % of the average responses consisted of the precise generalization, 'transportation'. The remaining 17 % superior and 58 % average responses were more specific ('motors in both'); or vaguer ('both go'); or clearly irrelevant ('people made both').

Table 7. *Definitions of 'puddle'*

Definition	Frequencies		
	Observed*		Expected
	Av.	Sup.	
A. Precise synonym ('standing water')	6	14	10.0
B. Descriptive explanation ('a place where rain makes mud in the road')	9	18	13.5
C. Vague characterization; situational illustration ('some water'; 'puddles are in back of my house after rain')	6	1	11.5
D. Repetitive or inadequate phrases ('a puddle of water'; 'it rains')	14	2	
Totals	35	35	35.0

* $\chi^2 = 18.766$, $P < 0.001$ (2 D.F.: $P < 0.001$, $\chi^2 = 13.82$).

IV. DISCUSSION AND CONCLUSIONS

It should be stressed that the limitations in scope of this brief study demand tentative rather than definitive conclusions. Among other things, there are (as yet) no reliability figures. Inter-observer reliability was very high (0.997) in an early study which utilized the qualitative approach (Terman & Merrill, 1937, p. 303, n.), and hence might be so, by analogy, in the present study, especially since this investigator was restricted to non-judgmental recording of behaviour. The Vigotsky presents the greatest barrier to an estimate of task reliability, in terms of parallel forms (some tasks bear simple retesting; the split-half technique appears less appropriate to a heterogeneous battery), but one study reports some progress in this direction (Semeonoff & Laird, 1952), and a gratifying stability of individual behaviour patterns.

It also should be kept in mind that the investigator both selected and tested the subjects. In order to prohibit any subliminal 'halo' effect, all identifying data for the subjects were filed away 3 months prior to testing, and protocols were coded and well-shuffled, after testing, with another 3 months' hiatus before categorization. Furthermore,

the rigid adherence to recording and sorting in terms of concrete referents virtually eliminated any opportunity for subjective assessment.

On the other hand, this latter restriction also limited the scope of the investigation. The extension of observations and analysis through the use of observer-teams, as well as recording devices, would permit study of endless nuances of behaviour glossed over in this initial survey. In the 'Word-naming' task, for instance, one could explore differential trends in a more subtle, *overall* coherence of thought processes, and in the variability, extensiveness, and generality of content categories. In the Vigotsky tasks, initial reactions call for further study, along with the flexibility and relevance of presolution hypotheses, and working attitudes (labile or intensive concentration; preference for or avoidance of overt manipulation; self-sufficiency, etc.). In the Rorschach situation, the study of differential organizations of a *constant* area may be extended to major details, as well as wholes, with the elaboration and precision of a response (*within* the $F+$ to $F-$ dichotomy) calling for consideration, among other qualities.

Meanwhile, certain tentative conclusions, relevant to this particular procedure and normative group, may be offered. (1) The mental processes of superior children were significantly more coherent, as well as more vigorous or productive, than those of average children. (2) Subtle behaviour reactions to an elementary task significantly differentiated intellectual levels, while grosser task achievements did not. (3) Superior children significantly more often abstracted and named conceptual principles, while average children remained 'situation-bound'. (4) Superior children demonstrated a significantly lesser degree of conceptual rigidity than average children. (5) Superior children required significantly less structure than average (and persevered significantly longer without aid) in a complex conceptual problem, and they made fuller use of assistance, when given. (6) Significantly more superior than average children perceived and manipulated multiple concepts. (7) Superior children had a significantly firmer grasp than average on demonstrated conceptual principles. (8) Logical incongruities were spontaneously recognized significantly more often by superior than by average youngsters. (9) Superior children were significantly more alert or attentive to camouflaged omissions and discrepancies. (10) Average children showed a significant restrictive trend towards 'global' percepts, while superior youngsters added or preferred differentiated percepts. (11) Superior children tended to define even the most familiar and mundane words with synonyms or comprehensive descriptions, in significant contrast to average indefiniteness or over-specificity.

Apart from the specific findings, it is hoped that this paper accomplishes its broadest purpose, of stimulating a novel orientation toward the analysis and subsequent measurement of the more fundamental intellectual processes. The battery itself, of course, may be revised and extended in an endless variety of ways, and applied to many other subject-groups, both normal and abnormal, toward the gradual accretion of a body of knowledge—a set of qualitative norms—of differential intellectual processes, at various life-stages, and toward truly fertile, experimentally based, dimensional quantification of intelligence.

This paper is based on a thesis offered in partial fulfilment of the requirements for the M.A. degree at the University of Florida, June 1955. The author wishes to express her appreciation to Dr J. C. Dixon and Dr R. J. Anderson for their unflinching interest and stimulating criticisms.

REFERENCES

- AMES, L. B., LEARNED, J., METRAUX, R. W. & WALKER, R. N. (1952). *Child Rorschach Responses: Developmental Trends from Two to Ten Years*. New York: Hoeber.
- ANASTASI, A. & FOLEY, J. P. (1954). *Differential Psychology*. (Rev. ed.) New York: Macmillan.
- BINET, A. & SIMON, T. (1916). *The Development of Intelligence in Children*. Baltimore: Williams and Wilkins.
- FEIFEL, H. & LORGE, I. (1950). Qualitative differences in the vocabulary responses of children. *J. educ. Psychol.* **41**, 1-81.
- HANFMANN, E. & KASANIN, J. (1937). A method for the study of concept-formation. *J. Psychol.* **3**, 521-40.
- HEARNSHAW, L. S. (1951). Exploring the intellect. *Brit. J. Psychol.* **42**, 315-21.
- HEIDBREDER, E. F. (1948). Studying human thinking. In T. G. Andrews (ed.), *Methods of Psychology*. New York: Wiley.
- HERTZ, M. R. (1951). *Frequency Tables for Scoring Responses to the Rorschach Ink Blot Test*. Cleveland: Western Reserve University Press.
- HERTZ, M. R. & EBERT, E. H. (1944). The mental procedure of 6 and 8 year old children as revealed by the Rorschach ink-blot method. *Rorschach Res. Exch.* **8**, 10-30.
- MCCOLSKEY, A. S. (1955). *An Investigation of Qualitative Differences in Intelligence between Average and Superior Children*. Unpublished Master's Thesis, University of Florida.
- MEYER, G. & THOMPSON, J. (1952). The performance of kindergarten children on the Rorschach test: a normative study. *J. Proj. Tech.* **16**, 86-111.
- RAY, J. J. (1936). The generalizing ability of dull, bright and superior children. *Peabody Contr. Educ.* No. 175.
- RICHMOND, W. K. (1953). Educational measurement: its scope and limitations. A critique. *Brit. J. Psychol.* **44**, 221-31.
- SEMEONOFF, B. & LAIRD, H. J. (1952). The Vigotsky test as a measure of intelligence. *Brit. J. Psychol.* **43**, 94-102.
- STAVRIANOS, B. (1942). An investigation of sex differences in children as revealed by the Rorschach method. *Rorschach Res. Exch.* **6**, 168-75.
- TERMAN, L. M. & MERRILL, M. A. (1937). *Measuring Intelligence*. Boston: Houghton Mifflin.
- WECHSLER, D. (1944). *The Measurement of Adult Intelligence*. (3rd ed.) Baltimore: Williams and Wilkins.
- WECHSLER, D. (1949). *Wechsler Intelligence Scale for Children*. New York: Psychol. Corp.

(Manuscript received 25 September 1956)

THE SOLVING OF THREE-TERM SERIES PROBLEMS

By IAN M. L. HUNTER

Psychology Department, Edinburgh University

Sixteen problems of the three-term series type are presented to each of 64 11-year-old children and 32 16-year-olds. The older children show evidence of solving these problems by articulated sequences of analytical judgments which accord with the varying logical structures of the problems, show no influence of atmosphere effect, show marked practice effects in progressing from one problem to the next, and solve problems involving the relatively abstract 'happier-sadder' relation as quickly as identically structured problems involving the relatively concrete 'taller-shorter' relation. In contrast, the younger children show but slight evidence of being influenced differentially by variations in logical structure, are susceptible to atmosphere effect, show smaller practice effects, and solve 'taller-shorter' problems more rapidly than 'happier-sadder' problems. The results are interpreted as demonstrating that increasing age brings increasing appreciation of the structural characteristics of series as such, together with increasing skill in dealing with serial relations which are progressively more remote from the perceptual-motor level of behaviour.

I. INTRODUCTION

'Tom is taller than Dick: John is shorter than Dick. Which of these three boys is the tallest?' This exemplifies what is here called a three-term series problem. Consider the series $A > B > C$. This series may be transformed into a set of two premises, each premise comprising two terms and a serial relation. (For a discussion of the logical properties of series see, for example, Stebbing, 1933, pp. 172ff.) When such a set of premises is presented to someone who is required to derive a three-term series from it, then the person is said to have a three-term series problem. This class of problem is not completely representative of series problems at large since it is restricted on three counts. First, the items to be seriated are not presented in concrete actuality as, for example, in Piaget's experiments with sticks, dolls, etc. (1952, pp. 96-157); rather they are presented verbally. Secondly, the series to be derived comprises the minimum possible number of items, namely, three; this contrasts with the four- and five-term series employed, for example, by Pinard, Barbeau, Laurendeau & Parant (1954) in their adult intelligence test. And thirdly, the problems, as stated, make no use either of the 'between' relation or of negative statements such as ' A is not taller than B '. Nevertheless, the very restrictedness of this class of problems makes for easier experimental investigation of at least some of those psychological processes which would seem to be fundamental to the solving of any type of series problem. The central question is: by what psychological processes do people derive the series from the given set of premises? Some part of an answer has been offered by Piaget (1921, 1928) but by far the greatest amount of data bearing on this question is furnished by Burt (1919). The discussions of Burt, with his predominantly statistical approach, and of Piaget, with his predominantly 'clinical' approach, suggest a number of variables likely to affect the solving of three-term series problems. It is the purpose of the experiment reported below to examine some of these variables further. But before describing this experiment, it may be best to clarify some points of terminology and present a theoretical outline of the solving process.

(i) *The structure of three-term series problems*

Consider the statement that A is taller than B who is taller than C . This series may be represented symbolically as $A > B > C$ or as $C < B < A$. It may also be presented as a set of two premises, the four basically different premise structures being as follows: I, $A > B : B > C$; II, $A > B : C < B$; III, $B < A : C < B$; IV, $B < A : B > C$. A , B , and C are the *terms* of the premises and likewise of the series. Any given serial relation is represented by $>$ (or by $<$) while $<$ (or $>$) represents the same relation *reversed*. The *converse* of any premise is that proposition in which the terms are transposed and the relation reversed, e.g. $B > C$ is the converse of $C < B$. The *referent* of any premise is its first term and the *referee* is its second term, e.g. A is the referent of $A > B$ and the referee of $B < A$. Two premises are *linked* where they have one term in common, e.g. $A > B : B > C$ are linked but $A > B : C > D$ are not. Two linked premises are *isotropic* when each contains the same unreversed relation and also when the common term is the referee of the first-encountered premise and the referent of the second premise, e.g. $A > B : B > C$ and $C < B : B < A$. When the relation of one linked premise is the reverse of that contained in the other, then these premises are *heterotropic* e.g. $A > B : C < B$. Likewise, when the relations are identical but the common term is the referent of the first premise and the referee of the second, then the linked premises are, again, heterotropic, e.g. $B < A : C < B$. (Notice that, since only two premises are involved, they will necessarily be encountered consecutively. With problems involving series of four or more terms, linked premises may be non-consecutive, e.g. $A > B : C > D : B > C$. In such a case, the person must, after encountering the first premise, suspend judgment, and either wait or search for that second premise which is linked to it. Examples of problems made more difficult by the non-consecutive presentation of linked premises are to be found in the relational sorites of Lewis Carroll (1895), and in the series problems of Pinard *et al.* (1954).)

Table 1

	First set	Second set
I, 1	$A > B : B > C > ?$	$C < B : B < A < ?$
I, 2	$A > B : B > C < ?$	$C < B : B < A > ?$
II, 1	$A > B : C < B > ?$	$C < B : A > B < ?$
II, 2	$A > B : C < B < ?$	$C < B : A > B > ?$
III, 1	$B < A : C < B > ?$	$B > C : A > B < ?$
III, 2	$B < A : C < B < ?$	$B > C : A > B > ?$
IV, 1	$B < A : B > C > ?$	$B > C : B < A < ?$
IV, 2	$B < A : B > C < ?$	$B > C : B < A > ?$

The premises having been presented to the person, he may be required to derive the series by being asked one of several different questions. In the present study, only one form of question is employed, namely, 'Who is tallest/shortest?' In other words, the question requires the person to state that term which stands to the other two in the asymmetrical relation expressed by the superlative adjective (e.g. is tallest of) instead of by the comparative (e.g. is taller than). Thus each of the above four premise structures may be followed by one of two questions, that is, either by (1) 'Who is tallest?' which may be symbolized as ' $> ?$ ', or by (2) 'Who is shortest?' which may be symbolized as ' $< ?$ '. The eight different problem structures which result are shown in the first column of Table 1 preceded by a 'code reference' in terms of which each problem structure will be

referred to hereafter. The second column shows a further eight structures which are identical with the first eight except that, in each structure, there is a reversal of all the relations involved and a corresponding change in the locations of the terms. Since the ensuing discussion applies equally well to both alternative sets of eight structures, reference will be made only to the structures listed in the first column. (A complete study of three-term series problems involves considering a total of sixteen different structures and these may be derived by arguments other than that outlined above. However, as Table 1 shows, half of these structures are, as it were, psychologically equivalent to the other half. Starting with the series $C > B > A$ instead of $A > B > C$, a further set of sixteen structures may be derived which are, in effect, identical with those shown in Table 1.)

(ii) *Reorganizings involved in deriving the series*

The covert psychological performances presumed to be involved in deriving a series from a set of two-term premises are here inferred on the basis of two assumptions. The first is that seriating necessitates the successive judging of premises which are not only linked (this is a logically necessary condition) but also isotropic. That is, they must be encountered in that sequence which James (1891), in discussing his 'axiom of skipped intermediaries', described by saying that 'any number of intermediaries may be expunged without obliging us to alter anything in what remains written' (p. 646). If two linked premises are not isotropic, the thinker must therefore make them isotropic by reorganizing one or other of the premises. The second assumption is that, where reorganizing is required, it is the second-encountered premise which tends to be reorganized rather than the first. That is, the first premise establishes the direction of the series, and predisposes the thinker to make the second premise isotropic with the first. Now, as regards the reorganizing performance required to make the two linked premises isotropic, this may (in optimally efficient reasoning) take one of three distinguishable forms according as to whether the problem is of the structure designated above as II, III, or IV. With II ($A > B : C < B$), the second premise requires simple *converting* to give $A > B : B > C$. With III ($B < A : C < B$), the second premise must be 'brought forward' or *reordered* to give $C < B : B < A$. With IV ($B < A : B > C$), isotropy may be achieved by two alternative procedures. In one of these, the second premise is both converted and reordered. Thus, $B < A : B > C$ becomes $B < A : C < B$ and then $C < B : B < A$. In the alternative procedure, it is the first-encountered premise which is converted. Thus, $B < A : B > C$ becomes $A > B : B > C$. But in order to achieve this converting, it is necessary to back-track on the first premise and reverse the direction established by this premise. Both of these alternative procedures involve, therefore, converting as well as some additional reorganizing. In the *converting-reordering* performance, it is the reordering which is additional. And in the *reversing-converting* performance, there is an additional reversing, that is, a 'going back' to and a reconsidering of the already encountered first premise and a reversing of the direction which has already been established by it. It is of interest that this reversing process has been singled out by Piaget in his discussions of 'operations', and by animal psychologists in discussing the 'reverse detour' as 'a back-and-forth movement inconsistent with one's natural expectation of a straightforward progress toward the goal' (Woodworth & Schlosberg, 1954, p. 834).

iii) *Atmosphere effects*

The above discussion treats the deriving of a three-term series as a sequence of analytical judgments of the relation and identity between successively encountered terms. But both Burt and Piaget found that, with children, the deriving may be strongly influenced by the 'general feel' or 'global impression' of the problem: that certain forms of problem predispose the thinker to accept that particular answer which 'looks right' although it may or may not be correct. This influence of 'global impression' has received its most unambiguous confirmation in the study of adult syllogistic reasoning and has been called, by Woodworth & Sells (1935), the 'atmosphere effect'. Another and particularly compelling instance of an 'atmosphere effect' which derives from the structural characteristics of the problem has been reported recently by Hunter (1957).

From Burt's discussion of three-term series problems, it would appear that atmosphere may derive from two closely related sources. These may be called 'direct statement' and 'inclusiveness', respectively. The effect of 'direct statement' is described by Burt as follows. 'With the statement "Jack runs slower..." ringing in the memory, a child asked "who is the slowest?" naturally tends to say "Jack runs slowest"' (p. 126). This source of suggestion will presumably operate most strongly in structure II where one premise contains the given relation and the other the reversed relation and where there is no conflict between the two direct statements. Consider the problem '*A* is taller than *B*: *C* is shorter than *B*. Who is tallest?' Here, the answer suggested by direct statement is the referent of the first premise, i.e. *A*. In a sense, the thinker is actually told that *A* is the tallest. Likewise, the referent of the second premise is suggested in answer to the question 'Who is shortest?' Therefore, irrespective of the question asked, problems of structure II suggest the correct answer by 'direct statement'. In contrast, problems of structure IV would suggest, by 'direct statement', the wrong answer, namely, the middle term. In this case, however, the two direct statements conflict, *B* being taller in one premise and shorter in the other. In consequence, it is doubtful whether atmosphere effect deriving from 'direct statement' would operate in structure IV.

'Inclusiveness' has been discussed by Piaget (1921) in relation to the following problem of structure IV, 2. 'Edith is fairer than Suzanne: Edith is darker than Lili. Which is the darkest, Edith, Suzanne, or Lili?' As Burt points out, the statement 'Edith is darker than Lili' suggests that both Edith and Lili belong to the dark group, Edith being merely more intensely marked by the special character of that group; hence, the child is apt to select Lili as the dark one when asked if she is darker than Suzanne. In Piaget's words, it is as though the child reasoned as follows. 'Edith is fairer than Suzanne, so they are both fair; Edith is darker than Lili, so they are both dark: therefore Lili is dark, Suzanne is fair, and Edith is between the two. In other words, owing to the interplay of the relations involved in the test, the child, by substituting the judgment of membership (Edith and Suzanne are "fair" etc.) for the judgment of relation (Edith is "fairer than" Suzanne), comes to a conclusion which is exactly the opposite of [the logical conclusion]' (1928, pp. 87-8). (It should be mentioned that the conclusions drawn by Piaget from his 1921 study are somewhat vitiated not only by his 'clinical method' but also by his use, in rephrasing the problems, of the words *blonde* and *brune* alongside the more relational words *claire* and *foncée*.) In the problem structures outlined above, 'inclusiveness' would favour an incorrect answer with IV and a correct answer with II.

In summary, if the two atmosphere effects were to operate, structure II would be easier because of both 'direct statement' and 'inclusiveness' while structure IV would be more difficult to solve because of 'inclusiveness' but not, presumably, of 'direct statement'.

(iv) *Reorganizing and ease of solving*

The processes suggested above as operating in the solving of three-term series problems are, at best, directly observable only to a limited extent through retrospecting and, at worst, not amenable at all to direct observation. However, if the ease with which any particular problem is solved were a function of the number and complexity of the reorganizings required, then these reorganizings would manifest themselves in the relative difficulties of each problem structure. Problems containing heterotropic linked premises would be more difficult than problems of structure I where the premises are already isotropic: IV would be more difficult than problems requiring only converting (II) or reordering (III): and if atmosphere effects operated, the negative atmosphere problems (IV) would be more difficult than the positive (II). Such differences in relative difficulty would reveal themselves either in the frequency with which the problems were solved under standard conditions or in the time required for solving, it being supposed that every additional reorganizing process increases both the likelihood of error and the time required to arrive at the correct solution.

II. PROCEDURE

(i) *Subjects*

There were two groups of subjects. The first comprised 64 children (32 boys and 32 girls) aged between 10 years 5 months and 11 years 11 months, with a mean age of 11 years 2 months. These children were drawn from two classes of the Moray House Demonstration School, Edinburgh, and were, as a group, above average intelligence. The second group comprised 32 children (14 boys and 18 girls) aged between 15 years 1 month and 16 years 11 months, with a mean age of 16 years 1 month. These children were also above average intelligence, being prospective candidates for the Scottish School Leaving Certificate. They all attended Portree Senior Secondary School, Isle of Skye. The reason for having twice the number of 11-year-old subjects will become apparent when the results are considered. Briefly, such small inter-structure differences emerged from the first 32 11-year-olds studied that it was thought advisable to repeat the whole experiment with a further group of 32 subjects.

(ii) *The problems used*

The sixteen problems used are shown schematically in Table 2, numbered 1 to 16. In the third column of that table, the structure of each problem is indicated by the 'code reference' described in the Introduction. The exact wordings of problem nos. 1, 2, and 3 were as follows. 'George is taller than Harry: Harry is taller than Willie. Which of these three boys is the tallest?' 'Lily is sadder than Olive: Lily is happier than Edith. Which of these three girls is the saddest?' 'Tom is taller than Dick: John is shorter than Dick. Which of these three boys is the tallest?' Each succeeding problem was worded in exactly the same way, the terms in each being either three short and familiar girls' names or three short and familiar boys' names. The specific names used were varied from problem to problem to avoid confusion; and the problems containing girls' names alternated with those containing boys' names. Throughout the sixteen problems, each of the eight different structures listed in the first column of Table 1 occurred twice, once with the 'taller-shorter' relation (these problems are designated by asterisks in Table 2) and once with an alternative serial relation. For the 16-year-olds and half of the 11-year-olds, this alternative was the 'happier-sadder' relation. 'Happier than' was substituted for the symbol '>' and 'sadder than' for the symbol '<'. For the remaining half of the 11-year-olds, 'warmer-colder' was the alternative relation: 'warmer than' was substituted for '>' and 'colder than' for '<'. Thus each of the 32 16-year-olds

tackled eight problems involving the 'taller-shorter' relation interspersed among eight problems, of identical structure, involving the 'happier-sadder' relation. Each of 32 11-year-olds tackled exactly the same problems as the members of the older group. And each of the remaining 32 11-year-olds tackled sixteen problems of which eight involved the 'taller-shorter' and eight the 'warmer-colder' relation. Each problem was typed on a separate white 6 in. by 4 in. card. The first premise occupied the first line, the second premise a second line directly beneath the first, and the question a third line beneath the other two.

ii) *Administration*

The procedure for both groups of children was identical. The children were tested individually in a room of the school attended by them. Each was given twenty problems in all, that is, four practice problems plus the sixteen test problems. When the child was seated and rapport established, he was instructed as follows. 'Now, we're going to do some simple problems and I think you'll find them easy enough. Here's the first one. Just read it over to yourself.' The first practice problem was handed to the child. 'Now, you see the question there at the end? Tell me what you think the answer is.' When this first problem had been correctly answered, the card was removed and the experimenter continued. Here's another problem much the same as before. But this time I'm going to time you with this watch to see how long you take. Tell me the answer as quickly as you can without, of course, hurrying so that you make mistakes. And if you do make a mistake, I'll tell you so that you can have another try.' With this, the experimenter presented the second practice problem, and started a stop-watch when the child began to read the problem. If a wrong answer was given, the experimenter said 'No, try again', and recorded the wrong answer along with the time at which it was given. The final (and usually the only) time recorded was that at which the correct answer was given. These times were entered in a prepared record blank to the nearest second; and the timing was carried out as unobtrusively as possible, without attempt to conceal from the child the fact of his being timed. Testing continued in this manner with the remaining practice problems and the sixteen test problems. The four practice problems were taken from Kurt's (1919) graded series. They were tests nos. 2, 3, 8, and 11 of that series and were always presented in that order. The test problems were presented to the first sixteen children in the same cyclical order (i.e. nos. 1, 2, ..., 16) but each child began with a different problem. Likewise, each of the second sixteen subjects began with a different problem, the problems now being in the constant cyclical order of nos. 16, 15, ..., 1. Taking the group as a whole, each problem was thus tackled first, second, third, and so on by an equal number of children with the effect of counterbalancing the influences of possible practice effects on any particular problem. At no point was the child questioned about his 'reasons' for giving any particular answer. Occasionally, however, the child would spontaneously offer some explanatory remark; and, when this happened, the remark was recorded.

iv) *Treatment of results*

Because, on the average, the 16-year-olds solved the problems more quickly than the 11-year-olds, the results for each group had to be treated separately. The general treatment of results was, however, identical for both groups. The basic assumption underlying the design of this study was that the introduction of an additional 'step' in the thinker's thought processes would involve an increase in the time taken to arrive at the correct solution. And the question was: did this problem take relatively longer to solve than that? For each problem there was, in each group, a distribution of either 64 or 32 time scores and the interest lay in comparing these distributions as between one problem and another. This comparison could not be made by *t*-tests because the distributions were not only skewed but also discontinuous as a consequence of some children having given a wrong answer before reconsidering the problem at the experimenter's prompting and arriving at the correct solution. The use of χ^2 was, however, appropriate. Accordingly, the median time for solving was found. This was between 11 and 12 sec. for the younger group and between 8 and 9 sec. for the older group. The total was then found of the number of instances in which a particular problem was solved correctly in less than the median time, and, likewise, of the number of instances in which that problem was not solved correctly in under the median time. However, it was discovered that this procedure threw away some of the data. Thus, one subject might have solved all sixteen problems in less than (or more than) the median time for the group; so, whatever time differences he showed between the solving of one problem and the next would not enter into the final results. The following modified procedure was, therefore, adopted. Instead of selecting the median time score for the total group of 32 or 64 children, each child was made to provide his own median value. This median

varied from child to child but, for each individual, there were eight problems falling below and eight above (allowance being made for ties between problems). The time scores of each child were thus treated separately, and a note made of those problems which were correctly solved in the eight shortest as contrasted with the eight longest times. The total was then found, for all the children in the group, of the number of instances in which each problem was (a) one of the eight easiest to solve, and (b) one of the eight most difficult. (One point about this method merits mention. It is the treatment of those instances in which a problem was correctly solved in less than the median time even after a wrong answer had been given and corrected. [These solutions were regarded as falling below the median time.] This treatment is admittedly questionable; but in fact it was of no practical consequence since such ambiguous instances occurred in only 5 out of the total of 1536 instances.) The results were worked out by both the above methods, that is, by taking the group median and the individual medians. Exactly the same interproblem differences emerged from each method; but the latter procedure proved its merit by revealing these differences in slightly enhanced form. It is the totals obtained by the latter (individual medians) method which are shown in Table 2 to two decimal places.

Table 2

No.	Structure	Ref.	11-year-olds, N=64. No. of solutions		16-year-olds, N=32. No. of solutions		
			Below median	Above median	Below median	Above median	
1*	$A > B : B > C$	$> ?$	I, 1	37.02	26.98	20.49	11.51
2	$B < A : B > C$	$< ?$	IV, 2	22.12	41.88	6.32	25.68
3*	$A > B : C < B$	$> ?$	II, 1	44.00	20.00	19.10	12.90
4	$B < A : C < B$	$< ?$	III, 2	34.35	29.65	20.57	11.43
5*	$B < A : B > C$	$> ?$	IV, 1	33.20	30.80	14.30	17.70
6	$A > B : B > C$	$< ?$	I, 2	28.83	35.17	24.80	7.20
7*	$B < A : C < B$	$> ?$	III, 1	34.58	29.42	15.25	16.75
8	$A > B : C < B$	$< ?$	II, 2	34.47	29.53	17.30	14.70
9	$A > B : B > C$	$> ?$	I, 1	33.02	30.98	25.62	6.38
10*	$B < A : B > C$	$< ?$	IV, 2	25.35	38.65	5.05	26.95
11	$A > B : C < B$	$> ?$	II, 1	33.10	30.90	15.01	16.99
12*	$B < A : C < B$	$< ?$	III, 2	30.10	33.90	14.12	17.88
13	$B < A : B > C$	$> ?$	IV, 1	24.35	39.65	7.93	24.07
14*	$A > B : B > C$	$< ?$	I, 2	34.43	29.57	20.06	11.94
15	$B < A : C < B$	$> ?$	III, 1	19.00	45.00	5.14	26.86
16*	$A > B : C < B$	$< ?$	II, 2	45.33	18.67	23.43	8.57

* Denotes those problems which involve the 'taller-shorter' relation.

III. RESULTS

(i) *Effect of practice*

People who perform a succession of similar tasks usually show progressive improvement from one task to the next. Do the children in this experiment show such an improvement? This question is answered by comparing the times taken to solve those problems which are tackled early in the sequence of sixteen with the times for those tackled late. With the 11-year-olds, the first four problems are solved in less than the median time in 115 (out of 256) instances, while the last four are similarly solved in 147 instances. The difference between these two proportions is significant at the 0.01 level ($\chi^2 = 8.005$). With the 16-year-olds, the first four problems are solved in less than the median time in 58 (out of 128) instances and the four problems tackled last in 88 instances. This difference is significant at well beyond the 0.01 level ($\chi^2 = 14.346$). Especially in view of the fact that there is only half the number of subjects in the older group, this difference is significant at a level which so far exceeds 0.01 as to lead to the conclusion that both groups show a practice effect, the older children much more than the younger.

The same conclusion emerges when the first and last two or three problems—or even the first and last problems only—are compared.)

ii) *Effect of premise structure*

The sixteen test problems contain only four different premise structures. These are shown in Table 3, together with the appropriate totals (to the nearest whole number) for both groups of children. The table also shows the χ^2 values between the proportions of above- and below-median instances for each structure and the levels at which they may be accepted as significant. Only three probability values are used, namely, 0.01, 0.02, and 0.05; and χ^2 values are corrected for continuity where the number of instances in any cell of the fourfold table is less than 50.

Consider, first, the results for the 16-year-olds. Structure II, which requires converting, is more difficult than I which is isotropic to start with. III, which requires reordering, is more difficult than I. III is also more difficult than II, which suggests that reordering requires more time for its performance than converting. Lastly, IV is the most difficult of all. These findings accord with what was said above concerning the reorganizings presumed to occur in the solving of three-term series problems.

Table 3

11-year-olds ($N=64$)					16-year-olds ($N=32$)				
Premise structure	No. of solutions		χ^2	P	Premise structure	No. of solutions		χ^2	P
	Below median	Above median				Below median	Above median		
II	157	99	3.581	—	I	91	37	3.855	0.05
I	133	123			II	75	53		
III	113	143	0.511	—	III	55	73	6.252	0.02
IV	105	151			IV	34	94	6.888	0.01

Compared with the 16-year-olds, the results for the 11-year-olds have three noteworthy features. (1) There is a reversal of the relative difficulty of I and II. This is strong evidence for the operation of a positive atmosphere effect in the younger children's solving of II. (2) This reversal apart, the trends (if not the magnitudes) of the inter-structure differences are the same as for the older subjects. As before, IV is the most difficult: it is not more difficult than III to a statistically significant extent, but it is significantly more difficult than I ($\chi^2=6.157$) and than II. Confirmation of the difference between IV and I comes from results obtained by Burt in using his test no. 12. The two versions of this test were as follows. 'Edith is fairer than Olive: but she is darker than Lily. Who is darker—Olive or Lily?' 'Lily is fairer than Edith: Edith is fairer than Olive. Who is the fairest—Lily or Olive?' The first version conformed to structure IV and was solved by 46% of Burt's 8-year-olds; the second version conformed to structure I and was solved by 72% of the 8-year-olds. Another result of Burt's which agrees with the present findings occurs in connexion with his test no. 16. This test was as follows. 'Three boys are sitting in a row: Harry is to the left of Willie: George is to the left of Harry. Which boy is in the middle?' This problem was of type III and was solved by almost exactly 50% of the 8-year-olds studied. When the premises were reordered to make the problem of structure I,

it was now solved by 61% of 9-year-olds. This difference of 11% between I and III is small and probably falls short of statistical significance as does the corresponding difference in the data of the present experiment. But in both Burt's case and the writer's, the difference between I and III is in the same direction; this suggests that III is indeed slightly more difficult than I for the younger children. The inter-structure differences, then, may be taken as having the same directions as those shown by the 16-year-olds (excepting the difference between I and II). But these differences are in every case smaller, and in most cases not statistically significant—a finding which is all the more striking when it is remembered that the number of younger subjects was twice that of the older. This raises the third feature of the 11-year-olds' results. (3) The amount of inter-structure difference is much less than for the 16-year-olds. The same inter-structure pattern is present, but is, as it were, blurred. This 'blurring' of differences reflects the fairly general finding that the judgments of younger children are more variable than those of older children both within the individual, and, the effects of specialized training apart, between individuals (Hunter, 1954). But more specifically, this 'blurring' indicates that the problem solving activities of the younger children are less influenced by variations in the problem structure and that these children are, as will be argued below, less appreciative of the structural characteristics of the problem situation.

(iii) *Effect of question asked*

Each of the four premise structures is followed by one of two questions. Does this difference in question affect ease of solving? Examination of the data shows that it does so in two significant cases only, both occurring with the 16-year-olds. For these older subjects, III, 2 is easier than III, 1 at the 0.02 level ($\chi^2=6.248$); and IV, 1 easier than IV, 2, at the 0.05 level ($\chi^2=4.083$). Both differences seem amenable to an explanation which is derived in the first instance from the writer's own retrospecting, and which, considering the present data as a whole, seems the only plausible explanation. The writer noticed that reorganizing served to isolate or emphasize one particular term of the series. Consider the premise pair (III) '*B* is shorter than *A*: *C* is shorter than *B*'. In deriving a series from these premises, the writer was very much aware of finding that *C* interrupted the direction of the series and of bringing forward *C* as the shortest member of the series. This made it possible to answer straight away the question 'who is shortest?' whereas the question 'who is tallest?' required a further reconsidering of the two remaining terms to decide which was indeed the tallest. If this introspective evidence is generalized into a formal statement, it would be: the term which is contained only in that premise which has to be reorganized is isolated on the ground that it should come at one end or the other of the entire series. This generalization is consistent with the 16-year-olds requiring more time to solve III, 1 than III, 2. It is also consistent with the slight, but not significant ($\chi^2=2.355$), tendency for II, 1 to be more difficult than II, 2; and also with IV, 2 being more difficult than IV, 1 (provided the reorganizing involved in IV is of the reversing-converting rather than of the converting-reordering type, i.e. provided it is the first premise which is converted). If this explanation holds, it furnishes some evidence for problems of structure IV being solved by a reversing-converting procedure. Additional evidence in favour of this method of solving comes from a comment made by Burt in connexion with his test no. 12. Burt observed that, when the children were afterwards 'explaining their reasons' for their answers, nearly every child who had answered

correctly stated that he had converted the disparate (i.e. the first-encountered) premise. To return to the present study, it is noteworthy that, with the 11-year-olds, the question asked has the same qualitative effects on solving time as it does with the older children but, in the case of these younger children, these effects are (as the reader might, by now, expect) not statistically significant.

iv) *Effect of relation used*

Previous investigators have found that problems with familiar terms tend to be easier than structurally identical problems with unfamiliar terms (Wilkins, 1928; Brownell & Stretch, 1931; both as cited by Johnson, 1955, pp. 244 and 249). Here, it was hoped to demonstrate the effect of differences in the relations embodied in problems of identical formal structure. Two relations were sought such that both would be reasonably familiar to the subjects while one would lead to greater difficulty than the other in reorganizing the problem situation. There are, at the least, two respects in which a comparison involving height might be easier than one involving happiness. First, differences in height can be visualized more readily; and most children seem to be visualizers. Secondly, differences in height involve only one dimension or aspect while differences in happiness are multi-dimensional, i.e. one person may be happier than another in a variety of different ways. *A* may be happier than *B* because of better health and *B* may be happier than *C* because of a clearer conscience; and the fact that manifold aspects have to be considered before deciding that *A* is therefore happier than *C* would tend to hold up the thinker for a longer time than the simple comparison of height. In short, while there are doubtless many complex and obscure contingencies in the environment and in the organism which go to make one attribute more abstract than another, it may be said that happiness is a more abstract attribute than height. From these considerations, the first expectation is that, in the present experiment, the eight problems involving the 'taller-shorter' relation will be more rapidly solved than the eight structurally identical problems involving the 'happier-sadder' relation. A second expectation concerns differences between the two age groups. From a developmental point of view, it might be supposed that the same contingencies (whatever their number and nature) which make for the relative abstractness of an attribute would also retard the individual's acquisition of this attribute as a means of categorizing and ordering the events in his surrounding world. Accordingly, the second expectation is that, provided the appropriate age range is being studied, the younger children will find it relatively more difficult than will the older to handle problems involving the more abstract attribute. Results confirm both of the above expectations. The first 32 11-year-olds studied solve the 'taller-shorter' problems in less than the median time in 145.06 (out of 256) instances while the 'happier-sadder' problems are similarly solved in only 110.93 instances. The difference between these proportions is highly significant ($\chi^2=9.031$). With the 16-year-olds, the 'taller-shorter' problems are easier to a quite insignificant extent ($\chi^2=0.633$). Thus, the younger subjects find the relatively abstract 'happier-sadder' problems more difficult than the structurally identical but relatively concrete 'taller-shorter' problems while the older subjects do not.

With the study of a further group of 32 11-year-olds, the opportunity was taken to contrast the 'taller-shorter' relation with the 'warmer-colder' relation. Problems involving this 'warmer-colder' relation are solved in less than the median time in 118.31 out of 256) instances while the 'taller-shorter' problems are similarly solved in 138.95

instances. Although the difference between these two proportions is not negligible, it is not quite statistically significant ($\chi^2 = 3.446$). It would seem that reasoning about entities which differ in degrees of warmth is, for the 11-year-olds, intermediate in difficulty between reasoning about differences in happiness and differences in height.

(v) *Effect of atmosphere*

The only direct evidence for the effect of atmosphere is that the 11-year-olds find I easier than I. The results do not reveal whether or not atmosphere operates in the case of the 16-year-olds; but if it does, it is insufficient to make II easier than the logically more straightforward I. The most parsimonious conclusion is that, with regard to II, a positive atmosphere operates in the case of the 11-year-olds but not in the case of the older children. Another conclusion, equally tentative, might be drawn with regard to the results of the 11-year-olds. It is that little or no atmosphere operates in IV. This conclusion is based on two rather slender lines of evidence. First, in IV, 'inclusiveness' would operate to suggest one specific and incorrect answer: also, as argued in the Introduction, the only atmosphere likely to operate in IV is that which derives from 'inclusiveness'. It follows that the most frequently given wrong answer ought to be that suggested by 'inclusiveness'. But in fact, each of the two possible wrong answers occurs equally often—suggesting that these errors arise from confusion rather than from any systematic source. Secondly, if any negative atmosphere operated in IV, it might be expected that this structure would give rise to much more difficulty than it does in fact. Thus, what evidence there is hints at the operation of little or no negative atmosphere in IV. Now if this is so, it suggests yet another conclusion. If 'inclusiveness' does not operate in IV, it probably does not operate in II either; so the positive atmosphere of II would derive largely if not entirely from 'direct statement'. This is not to suggest that 'inclusiveness' never operates to facilitate and retard the solving of three-term problems; 'inclusiveness' might well have operated had, say, the subjects been a year or two younger.

IV. CONCLUSIONS

This study began by analysing the class of three-term series problems into eight different types or logical structures. Factors likely to affect the solving of problems of each structure were then discussed in the light of Burt's pioneer investigation. One set of factors involved an atmosphere effect in which the global impression of the problem suggests one particular answer which may or may not be correct. The second set of factors involved analytical judgments such as would entail processes of reorganizing, namely, reordering, converting, and reversing: an additional process, that of isolating, was added later as a by-product of the above reorganizing processes. Just as Bruner, Goodnow & Austin (1956) used reception and selection activities to construct 'ideal strategies' for the attainment of concepts so, in the present study, these reorganizing activities were used to construct the optimally efficient procedure for solving each problem of a given structure, that is, the minimal sequence of transformations which the given statement must undergo to yield an isotropic series and, hence, the solution. A variety of three-term series problems were then presented to groups of 11- and 16-year-olds on the assumption that the time taken to solve each problem would be a function of the factors mentioned above.

The main findings are summarized in the abstract at the beginning of this paper. But to these the following must be added. First, the older children require less absolute time

to solve the problems. Secondly, the older children find converting easier than reordering. Thirdly, there is some evidence that problems of structure IV are solved by reversing-converting rather than by converting-reordering. Fourthly, the atmosphere effect shown by the younger children appears to affect only the solving of problems of structure II and to derive from 'direct statement'. Lastly, as far as premise structure is concerned, the order of increasing difficulty is I, II, III, and IV for the 16-year-olds and II, I, III, and IV for the 11-year-olds.

The several findings of this study are self-consistent and seem to exemplify two developmental trends. The first is that increasing age brings increasing appreciation of the structural characteristics of series. The second is that increasing age brings increasing skill in handling progressively more abstract relations, that is, relations which are progressively more complex and more remote from the perceptual-motor level at which, as Piaget has shown, the performance of seriating is first accomplished by the developing child. The first of these two trends must now be discussed in more detail.

'When I try to work out what is involved in the step sequences of my experiments . . . there seems to me to be two outstanding constituents in the thought process.' One of these is 'the appreciation of the structural characters of the system within which thinking is being exercised'. It is in these words that Bartlett (1956, p. 449) echoes, to some extent, the conclusions of Wertheimer (1945) and alludes to a feature which would seem to be of the first importance in directed thinking. What is meant by 'appreciation of structural characteristics' may be illustrated by citing one of the present writer's as yet uncompleted studies. Students are presented with a succession of simple five-premise sorites such as the following. 'All boys with red hair have grey eyes: all tall boys have red hair: all thin boys work hard: all boys with grey eyes are thin: all boys who play football are tall.' The student is asked to draw that conclusion which follows when each and every one of the above statements is taken into account and to 'think aloud' while doing so. When the student first tackles such a problem, his activity is haphazard. He may combine a couple of premises, draw a part conclusion, leave it, combine a further pair of premises, draw a second part conclusion, and try to see if this can be combined with the first part conclusion. In short, he solves the problem by a sequence of steps which is erratic and subject to many chance factors. But after solving a few more problems of this type, his performance is characteristically transformed. He largely ignores the order in which the premises are presented: he reads through the statements in search of that one of the two terms ('retinends') which are not repeated and which he will take as his starting point: and from this starting point, he considers each premise in such a sequence that the terms form a consecutive chain with identical terms ('eliminands') juxtaposed. Not only does he solve the problem in a shorter time but he does so in a more systematic way for, once he has selected his initial 'retinend', the sequence of steps taken is fully predictable. At this stage, he has, in his own words, 'got the hang of the problem'; in Bartlett's words, he appreciates the problem's structural characteristics. The difference between his initial and later performance is comparable to the difference between the execution of a well-articulated and practiced perceptual-motor skill such as typewriting and the performance of that same skill by a less practised person with his lack of assurance and his hunt-and-peck methods. It is not a rigid sequence of part-activities which has been learned but a co-ordinated and flexible system of activities from which the person can readily select those appropriate to deal with particular circumstances.

To return to the three-term series problems, the difference between the younger and older children is that the latter approximate more closely to this ideal appreciation of structure, that they behave, in a manner of speaking, more logically by more nearly following the optimal procedure for solving each problem. They can more readily treat the given premises as components of a series and more readily perform whatever reorganizings may be necessary in order to derive the series most expeditiously from them: they more easily recognize, and during the course of the experiment come to recognize, each problem as representative of a class of problems; and they know better just how to proceed in progressing towards the solution. The older children, as contrasted with the younger, have, the writer would suggest, a better appreciation of the structural characteristics of series problems. This better appreciation manifests itself in their solving the problems more rapidly, in their behaving in closer accord with the logical demands of the problem structure, in their resisting the suggestiveness of diffuse global impressions, and in their increasing to a greater extent their problem-solving efficiency over successive problems even though the pattern of reorganizings varies from problem to problem. The younger subjects arrive at the solution by less economic and predictable reorganizings with the result that their solving-time is both longer and less affected by variations in the logical structure: and being less able to analyse out the essential structure of the problems, they are more influenced by such irrelevances as global impression and less able to react to similarities between successive problems.

The writer wishes to thank Mr Iain Murray for conducting the experiment with the 16-year-olds. Thanks are due to Mr Mowat, Miss Brown, Miss Primrose, and Mr Spears of Moray House School, Edinburgh, and to Mr Murray of Portree Senior Secondary School for their kindness in making both subjects and testing facilities available. The writer also wishes to thank Sir Cyril Burt for much detailed and highly constructive comment on earlier versions of this paper.

REFERENCES

- BARTLETT, F. C. (1956). Some experiments about thinking. *Proc. Roy. Soc. B*, **145**, 443-51.
- BRUNER, J. S., GOODNOW, J. J. & AUSTIN, G. A. (1956). *A Study of Thinking*. London: Chapman and Hall.
- BURT, C. (1919). The development of reasoning in school children. *J. Exp. Ped.* **5**, 68-77 and 121-7.
- CARROLL, L. (1895). *Symbolic Logic*. As selectively reproduced in *The Complete Works of Lewis Carroll*. London: Nonesuch. 1939.
- HUNTER, I. M. L. (1954). Tactile-kinaesthetic perception of straightness in blind and sighted humans. *Quart. J. Exp. Psychol.* **6**, 149-54.
- HUNTER, I. M. L. (1957). Note on an atmosphere effect in adult reasoning. *Quart. J. Exp. Psychol.* **9**, 175-6.
- JAMES, W. (1891). *The Principles of Psychology*. Vol. II. London: Macmillan.
- JOHNSON, D. M. (1955). *The Psychology of Thought and Judgment*. New York: Harper.
- PIAGET, J. (1921). Une forme verbal de la comparaison chez l'enfant. *Arch. Psychol.* **18**, 141-72.
- PIAGET, J. (1928). *Judgment and Reasoning in the Child*. London: Kegan Paul.
- PIAGET, J. (1952). *The Child's Conception of Number*. London: Routledge and Kegan Paul.
- PINARD, A., BARBEAU, G. L., LAURENDEAU, M. & PARANT, C. (1954). *Tests Différentiels d'Intelligence*. Université de Montréal.
- STEBBING, L. S. (1933). *A Modern Introduction to Logic*. London: Methuen.
- WERTHEIMER, M. (1945). *Productive Thinking*. New York: Harper.
- WOODWORTH, R. S., & SCHLOSBERG, H. (1954). *Experimental Psychology*. London: Methuen.
- WOODWORTH, R. S. & SELLS, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *J. Exp. Psychol.* **18**, 451-60.

(Manuscript received 27 October 1956)

THE INFLUENCE OF EXTRANEOUS STIMULATION ON THE REACTION TIME OF SCHIZOPHRENICS

BY J. TIZARD AND P. H. VENABLES

*Medical Research Council, Social Psychiatry Research Unit,
Institute of Psychiatry, London*

In a first experiment 74 chronic schizophrenics made 50 consecutive responses to the presentation of a stimulus light on each of two occasions. In an experimental run a 70 db white noise sounded between responses 21-30; in the control run there was no extraneous stimulation. Patients rated as 'withdrawn' showed an increase in speed of reaction in the presence of the noise, while there was no evidence to show that the extraneous stimulus affected those patients rated as 'sociable'. There was a general tendency for all schizophrenics regardless of classification to show a greater fall off of performance over time than normals. In a second experiment 30 sociable and 30 withdrawn patients responded to an auditory stimulus, in two different conditions of illumination. RT's of withdrawn schizophrenics were significantly faster when they worked in a brightly lit room than when they worked in a dimly lit room. Sociable patients showed a slight tendency to work faster in dim illumination. The findings are discussed.

I. INTRODUCTION

In a previous study (Venables & Tizard, 1956) it has been shown that the performance of non-paranoid, chronic schizophrenics on a continuous task differs from that of patients suffering from psychotic depression in that the schizophrenics show a more rapid decrement in performance, and greater reminiscence effects following a rest pause. In contrast to the schizophrenics, depressed patients show a typical 'reversal of the fatigue effect'. This finding confirmed earlier work by Kraepelin (1913) and by Hoch (1901) using different types of task, and related the phenomena to a single task. In the present experiments both paranoid and non-paranoid schizophrenics were tested in order to extend the scope of findings previously obtained, and to compare the two groups in their response to additional stimulation. Preliminary experimentation had indicated that those subjects who could clinically be thought of as withdrawn responded in a different way from those who appeared more extraverted. To put this clinical impression on a firmer basis a scale which was intended to measure withdrawal versus sociability was constructed.

The present experiments were initially based on the following assumptions: if, as suggested in our previous study, the reason that schizophrenics fall off rapidly in performance is because stimulus-responses connexions lead to an abnormally rapid build-up of internal inhibition (or I_R in Hullian terms, for at present it is not proposed to distinguish between the two), then a 'distracting' stimulus should have the effect of dissipating some of this inhibition, and should lead to an increment in performance. Such an effect has in fact been found by Mackworth (1950) using normal subjects, and the first experiment to be reported here showed that 'withdrawn' schizophrenics speeded up under the influence of a distracting stimulus. This experiment did not however show that the mechanism responsible for the phenomenon was that of disinhibition; and a second experiment, designed to minimize the possibility of facilitation occurring through disinhibition was therefore carried out to see if the effect could still be demonstrated. An

alternative hypothesis, that the effects were brought about by the 'tonic' effect of non-specific sensory stimulation on the cerebral cortex, was also proposed.

The aims of the experiments can be summarized as follows. (1) To compare the work decrement of paranoid and non-paranoid schizophrenics, and withdrawn as against sociable patients, on the reaction time task. (2) To test the theory that an extraneous stimulus, introduced into a reaction time series, will have a facilitating effect on the performance of schizophrenics. (3) To see whether schizophrenics rated as sociable (on the basis of a rating of social withdrawal by charge nurses or ward sisters) react differently under conditions of extraneous stimulation from those rated as withdrawn.

II. FIRST EXPERIMENT

(i) *Procedure*

Subjects were tested individually under standard conditions. They were required to respond as quickly as possible to the appearance of a visual stimulus by pressing down a toggle switch. A toggle switch rather than the conventional morse key was used, as its action was more clearly understood by seriously ill patients, and as the absolute speed of RT was not under consideration. The visual stimulus consisted of a ground glass screen 0.75 in. in diameter, illuminated from the rear, set in a 2 ft. by 2 ft. matt black panel. The intensity of the stimulus was 135 ft. candles, and the background illumination reaching the subject's eye was 1.7 ft. candles. A bell served as a warning signal preceding the appearance of the visual stimulus by 8.0 sec. The stimulus was presented for 3 sec. if no response was made, and the cycle of bell and light was repeated once every 15 sec. Each subject responded to 50 stimuli running consecutively; in the experimental series the extraneous stimulus sounded continuously during responses 21-30. The subjects were told that they might hear some 'static'—'like you get on the radio sometimes'—during the experiment, owing to 'interference', and this explanation proved sufficient for them to accept the situation naturally when the noise occurred.

Half the subjects performed first under control conditions, that is, they made 50 consecutive responses with no extraneous stimulus; the other half were tested first under experimental conditions, that is, where the additional extraneous stimulus sounded between responses 21-30. On a second occasion of testing the subjects worked under the condition which they had not previously experienced.

The extraneous stimulus was white noise produced by amplifying the discharge of a neon tube by means of a circuit having linear response from 20 to 18,000 c.p.s., and was reproduced by means of a single 12 in. diameter loud-speaker set in a 4 ft. by 4 ft. baffle situated 5 ft. from the subject's head. The intensity of the sound was adjusted to be 70 db as measured by means of a noise level meter. The intensity of both stimulus light and sound stimulus was kept constant by being supplied from a regulated voltage source.

Reaction times were measured by means of a cold cathode counter tube timer to the nearest millisecond.

(ii) *Subjects*

Forty-two male and 32 female subjects of whom 46 were classified as non-paranoid schizophrenics and 28 as paranoid schizophrenics were tested. All were less than 56 years of age and had been in hospital more than 1 year. The mean age of the male patients was 40.85 s.d. 8.94 years, and of the female patients 45.5 s.d. 9.19 years. No patient had undergone leucotomy, but some patients were under treatment by chlorpromazine, methopromazine, and similar drugs at the time of the investigation. (The effect of these treatments upon performance is described below.) Six male and 6 female nurses acted as normal controls. Their mean age was 26.16 s.d. 4.19 years.

Each patient was rated by two charge nurses or sisters on a scale constructed to measure sociability-withdrawal. There were in this scale 6 items, 5 of which were taken from the M.S.R.P.P. Scale developed by Lorr, Singer & Zobel (1951). (Items nos. 8, 52, 56, 57 and 58.) An additional item was included to give a measure of the number of friends the patient had in the ward. The average inter-rate reliability was $r=0.765$ for an N of 74. Reliabilities of 0.854 and 0.675 were obtained respectively for charge nurses rating male patients, and ward sisters rating female patients. Independence of the scale from age is shown by a correlation of ± 0.064 between the subject's score and age in years.

(iii) *Results*

For the purpose of the analysis, the 50 responses made by each subject were considered in 5 groups of 10 each, and further work was carried out on the mean values of each group of 10 responses. As the distribution of RT's and particularly of changes in RT under extraneous stimulation were not normally distributed, non-parametric statistics were employed. Because in this instance mean values of RT tended to obscure the picture owing to the abnormally large responses by some subjects, the results were analysed on the basis of change in direction of responses without taking note of their magnitude.

To examine the trends in performance of the schizophrenic subjects under the control conditions the performance of each subject during the control run was summarized by a linear regression line, and the slope constant examined. In 55 out of 74 subjects the slope was positive, indicating that RT became slower throughout; this as it stands is significant at the 0.01 level by the Sign Test (Dixon & Mood, 1946). In the case of the six normal controls, the slope was negative in five cases. Although the application of a χ^2 test in the case of such small numbers in two cells must be considered with caution, the result of 3.245 after applying Yates' correction is probably sufficient to indicate a significant association between the increase in RT and schizophrenia, as the size of the χ^2 gives probability value beyond the 0.02 level of significance.

Further examination, to determine whether the increase in RT during the control runs is associated more with paranoid than with non-paranoid schizophrenia, or is found more in withdrawn than in sociable schizophrenics, results in insignificant χ^2 of 0.016 and 0.621 respectively. The age difference of subjects showing positive or negative slope constants is not significant ($P > 0.05$). This lack of association with age confirms the previous finding (Venables & Tizard, 1956), where there was no significant association between decrement in performance on a repetitive task and age.

The test of the hypothesis that schizophrenic performance can be speeded up by an extraneous stimulus was made by examining the change in RT between the means of responses 11-20 and those of responses 21-30. The number of subjects whose mean RT's were greater during responses 21-30 than during responses 11-20, that is, who were slower while the extraneous stimulus was sounding, was 35, whereas the number of subjects who were faster while the extraneous stimulus was sounding was 39. The difference is not, of course, significant, and the second hypothesis—that an extraneous stimulus would facilitate the performance of schizophrenics—must be regarded as untenable.

The data were then examined to compare the performance of paranoid and non-paranoid subjects, and sociable and withdrawn subjects. Of the non-paranoid subjects, 29 were faster while the extraneous stimulus was sounding, while 17 were slower, and of the paranoid subjects 10 were faster, and 18 were slower. This association between diagnosis and change in response speed is significant as shown by a χ^2 of 5.21, $P < 0.02$.

To see whether 'social withdrawal' was a relevant factor in influencing change in response under an extraneous stimulus, patients were divided into two groups depending on whether they fell above or below the half-way point on the sociability rating. Of the withdrawn subjects, 23 were faster during the time the auditory stimulus was sounding, while 8 were slower. Of the sociable group 27 were slower, and 16 were faster. These figures gave a χ^2 of 9.89, which is significant at the 0.01 level. The difference between the

mean ages of subjects increasing and decreasing in RT under extraneous stimulation is not significant ($P > 0.05$).

In a normal control group who, for administrative reasons, were tested only once, two subjects increased and two decreased in RT while the auditory stimulus was sounding, while two remained the same. It is apparent that the auditory stimulus had little effect on the normal subjects.

As a check it is now desirable to examine the performance of the schizophrenic subjects under control conditions. The number of subjects showing a decrease in reaction latency from responses 11–20 to 21–30 is 28. The number showing an increase in latency is 46. If these subjects are now divided as before into paranoid and non-paranoid groups, among the paranoids 12 show a decrease in reaction latency, and 16 an increase; and of the non-paranoids 16 show a decrease in latency, and 30 an increase. That the change in latency under control conditions is not relevant to diagnosis is shown by an insignificant χ^2 of 0.504. A similar division into withdrawn and sociable groups shows by an insignificant χ^2 of 0.228 that there is no relationship here either.

The analysis so far has shown that both patients diagnosed as non-paranoid, and those rated as withdrawn, tend to decrease in RT (i.e. they were faster), while the extraneous stimulus was sounding; patients diagnosed as paranoid or rated as sociable increased in RT. However, it was noted that during the control run there was a tendency for all groups to increase in RT throughout the testing session. The data were therefore re-examined to see if the trend in performance during the extraneous stimulation differed from the trend shown by the same patients during the control run. A ranking technique was employed to make the comparisons.

If we consider the five periods of 10 responses, and assign a rank value to each on the basis of magnitude of RT we can compare the ranks obtained by each subject under the experimental conditions with the ranks of the same subjects under control conditions. The mean values of the ranks actually obtained are shown in Table 1.

Table 1. *Mean ranks of response time for five periods of 10 responses under control versus experimental conditions (Rank 1—fastest response)*

	Period					Total
	1	2	3	4	5	
Paranoid $N=28$						
Experimental	2.95	2.64	2.96	3.23	3.22	15.00
Control	2.96	2.68	2.66	2.57	3.13	15.00
Non-paranoid $N=46$						
Experimental	2.61	2.98	2.26	3.66	3.49	15.00
Control	2.34	2.63	2.92	3.51	3.60	15.00
Sociable $N=44$						
Experimental	2.57	2.61	2.92	3.61	3.28	15.00
Control	2.87	2.66	2.66	3.53	3.27	15.00
Withdrawn $N=30$						
Experimental	2.98	3.20	1.95	3.33	3.54	15.00
Control	2.13	2.63	3.07	3.53	3.64	15.00

It is seen from Table 1 that with the sociable and paranoid subjects the difference between the ranks obtained during Period 3 under control and experimental conditions is slight, being 0.26 and 0.30 respectively. These differences are not significant ($t=0.870$ for the sociable group and 0.816 for the paranoid group). We must conclude therefore

that any increase in RT is not due to the extraneous stimulation, but is part of the general pattern of response.

In the case of the withdrawn and non-paranoid subjects there is a greater difference between the ranks obtained during Period 3 under the two conditions ($C3 - E3 = 1.11$ and $.66$ respectively). These differences are significant in both cases; ($t = 3.42$, $P < 0.01$, and $= 2.26$, $P < 0.05$ respectively). We can therefore conclude that the effect of extraneous stimulation on these two groups of subjects is that of facilitation.

Owing to the small number of subjects (six), who were in the withdrawn paranoid group, further analysis of the data to investigate the interaction between the two factors of diagnostic category and social withdrawal yielded no significant results.

The responses of patients being treated with chlorpromazine and similar drugs were then examined. Of 23 such patients, 18 were rated as sociable, and only 5 as withdrawn. With these subjects the association between decrease in RT while the stimulus was sound and social withdrawal was weakened. χ^2 is reduced to 2.48 which does not reach an acceptable level of significance, probably because of the small number of withdrawn schizophrenics in this group.

iv) Discussion

The findings are as follows. Under control conditions performance of schizophrenics in this as in other experiments shows a significant 'fatigue effect', the extent of which is independent of diagnosis, age or 'social withdrawal'.

The effect of an extraneous stimulus was to facilitate the performance of non-paranoid and withdrawn schizophrenics. There was no evidence to show that a distracting stimulus affected the performance of sociable or paranoid patients.

The findings as they stand give partial confirmation of the prediction from the hypothesis that the RT's of schizophrenics can be speeded up through disinhibition in so far as one group of schizophrenics, whether by disinhibition or not, showed facilitation in their RT performance in the presence of a loud auditory stimulus. An alternative and perhaps more likely hypothesis is that in chronic, withdrawn schizophrenics the central excitatory state is lowered in some way. An increase in sensory input might be expected to have non-specific facilitatory effects on cortical and subcortical activities. Electrophysiological work summarized by Lindsley (1956) suggests that such effects can be found in animals, and work associated chiefly with the name of G. L. Freeman (e.g. 1948) has indicated that in normal human subjects excitability to various specific stimuli may be increased by an increase in background stimulation.

A second experiment was therefore carried out under conditions which reduced the possibility of facilitation occurring through disinhibition to a minimum. In this experiment RT's to an auditory stimulus were measured under two conditions of background illumination to see whether the additional non-specific sensory input coming from a brightly lit room would have facilitatory effects.

i) Procedure

III. SECOND EXPERIMENT

Forty-eight male and 12 female schizophrenics all less than 55 years old and whose mean age was 2.5 years S.D. 9.26 years, served as subjects. Thirty subjects were rated as 'sociable' using the procedure described earlier, while thirty were rated as 'withdrawn'. Twenty-one patients were being treated by chlorpromazine at the time of the experiment.

Each subject made 20 responses to a 1000 c.p.s. tone whose intensity was 60 db above threshold, presented through a loud-speaker under bright (18.0 ft. candles) illumination and 20 responses to a similar tone in a dim (0.3 ft. candles) illumination. Half the subjects worked first in bright and then in dim illumination, and half in dim and then bright illumination. The allocation of subjects was made randomly. The subject was allowed 3 min. in each condition of illumination to become accustomed to the situation before being tested. RT's were measured to the nearest millisecond by means of an electronic timer.

(ii) *Results*

Preliminary analysis showed that the order of illumination in which the subjects performed was not a significant variable; ($\chi^2=0.0026$, $P>0.90$). The first part of Table 2

Table 2. *Numbers of subjects giving faster RT's in dim or bright lights*

	Total group		Paranoid subjects		Subjects under drugs	
	Sociable	Withdrawn	Sociable	Withdrawn	Sociable	Withdrawn
Faster in dim light	19	9	3	0	8	0
Faster in bright light	11	21	1	3	5	8
χ^2		11.84		—		—
P		0.001		—		—

shows the number of subjects, divided on the basis of the sociability-withdrawal rating into 'withdrawn' or 'sociable' groups, who responded more quickly under bright or dim illumination. The association between rating and response pattern is highly significant ($\chi^2=11.84$, $P<0.001$). Of the sociable patients 19 were faster in dim illumination while 11 were faster in bright; and of the withdrawn patients nine were faster in dim illumination while 21 were faster in bright. As in the first experiment there is no statistically acceptable evidence that sociable patients are faster in one condition than in the other ($P>0.25$ using the Sign Test). The probability that the results obtained with the withdrawn patients is due to chance is however less than 0.05, using the Sign Test.

All but 7 of the patients were diagnosed as suffering from other than paranoid schizophrenia. Separate data from the seven paranoid subjects are shown in the middle part of Table 2, where it is seen that they behave according to the general pattern. The same tendency is shown by the 21 patients under drug treatment at the time of the experiment, whose data are given on the right hand side of Table 2.

(iii) *Discussion*

The results show that in an experimental situation in which the possibility of disinhibition as a mechanism for facilitation had been minimized, there was again a significant tendency for withdrawn patients to react more quickly when an extraneous stimulus was present. The alternative hypothesis, that an increase in sensory input has a non-specific facilitatory effect in withdrawn patients thus secures support.

This work was carried out in Springfield Hospital, London, by kind permission of the Medical Superintendent, Dr H. C. Beccle.

REFERENCES

- DIXON, W. J. & MOOD, A. M. (1946). The statistical sign test. *J. Amer. Statist. Ass.* **41**, 557-66.
- FREEMAN, G. L. (1948). *The Energetics of Human Behavior*. New York: Cornell University Press.
- HOCH, A. (1901). On certain studies with the ergograph. *J. Nerv. Ment. Dis.* **28**, 620-8.
- KRAEPELIN, E. (1913). *Psychiatrie* (vol. 3). Leipzig: Barth.
- LINDSLEY, D. B. (1956). Physiological Psychology. *Ann. Rev. Psychol.* **7**, 323-48.
- LORR, M., SINGER, M. & ZOBEL, H. (1951). Development of a record for the description of psychiatric patients. *Psychol. Serv. Cent. J.* **3**, no. 3.
- MACKWORTH, N. H. (1950). *Researches on the Measurement of Human Performance*. M.R.C. Special Report, no. 268. London: H.M.S.O.
- VENABLES, P. H. & TIZARD, J. (1956). The performance of functional psychotics on a repetitive task. *J. Abnorm. (Soc.) Psychol.* **53**, 23-6.

(Manuscript received 11 October 1956)

THE BOUNDARIES BETWEEN THE PERSON AND THE WORLD

A NOTE ON THE PAPER BY GARDNER MURPHY (*Brit. J. Psychol.* 1956, **47**, 88-94)BY LEON LITWINSKI, *Brussels*

Like Murphy I appreciate the field theories of Lewin and Angyal to which he refers in his paper (see Litwinski, 1951, 1956). My formulations, however, differ in some respects from Murphy's. In contrast to the widespread opinion that person and world can be sharply distinguished, Murphy emphasizes that one often melts into the other. In fact I made this same point in a paper presented to the Eleventh International Congress of Philosophy. There I quoted Ward (1951) who speaks of a pilot's feeling of one-ness with his machine. But, in my opinion, such cases must not be generalized too much. They belong rather to animistic states of mind which prevail in certain periods of development or in certain circumstances. They also belong to selfless experiences or impersonal states. They should be called identifications rather than fusions. Therefore, the notion that we tend to melt into our environment seems to be rather a metaphorical way of speaking. Like all analogies, they are safe if not pushed too far. The scientific statements require some rigorousness in the use of terms.

Under the command of love, for instance, a conjugal union may become so intimate that it may be called '*égoïsme à deux*'. Are we to conclude that the distinction between the two members of such union vanishes and that it is permissible to speak of osmotic fusion of two human beings? It seems that we must not go too far in formulating our theories in this way. And we must recognize a figurative mode of speech for what it is. 'If two persons of the same mind (*natura*) come together,' said Spinoza 'they form an individual twice as strong as either. . . Nothing, therefore, is more useful to man than man.' But the contrary may be argued: if two men are starving and there is only one loaf of bread, it is soon apparent that nothing is more detrimental to a man than another man. The man who finds that the woman he loves has another lover, makes a similar sad reflexion. Hence, in theoretical considerations there is constant need for stress on rivalry, on states of crisis or drama in human intercourse, resulting from the identity or similarity of individual needs and the too often limited means of satisfying them. We have to distinguish what the man seeks to get, what he can get and what he has got. Plotinus emphasized that man gets what he can and keeps what he has got. (*Lucrī quidquid est, id domum trahere oportet.*) The constant pressure of the outer and inner worlds tend to create for the individual a number of vital problems, of different magnitude, ranging from physical sufferings and discomforts to moral pains and uneasiness. So arises the search for cohesion, coexistence, adjustment, etc.

If it is true that, in the world of physics, man can conquer nature on condition that he obeys her laws, it is also true that, in social affairs, man can more freely integrate himself within a group if he respects the liberty of his fellow men. Here emerges the need to appreciate those limits that are functionally related to belongingness. Such a sense of limitation may be either underdeveloped or overdeveloped (another source of crisis extending from propensities to waste to those of avarice). By this sense of limitation however the individual may reveal wisdom or foresight. Recognition of limits may thus be therapeutic in education as well as in politics where stiffeners as well as softeners are required. The numerous material bonds, the various affectional or social ties are all relative in meaning. They may be strong or weak, lasting, ephemeral or transitory, developed or undeveloped. Their manifestations are observable not only in space but also in time. This is why, in my studies of personality, I stress the prospective character of bonds and ties between the subject and the 'object' considered in its extended sense, i.e. signifying not only material things.

To the fundamental question, are we one with the world or isolated from it, my reply is that we are neither one with the world nor separated from it. In varying degrees, the individual is prolonged in the outer world through his belongingness and he is projected in the future through his expectations. Thus, in my view, both the concept of belongingness and the notion of expectation are particularly important in the interpretation of the facts.

REFERENCES

- LITWINSKI, L. (1951). Toward the reinstatement of the concept of self. *Brit. J. Psychol.* **42**, 246-9.
 LITWINSKI, L. (1956). Belongingness as a unifying concept in personality investigation. *Acta Psychol.* **12**, 130-5.
 WARD, T. H. G. (1951). The psychological relationship between man and aircraft. *Brit. J. Med. Psychol.* **24**, 283-90.

(*Manuscript received 30 July 1956*)

PUBLICATIONS RECENTLY RECEIVED

Secondary School Selection. Ed. by P. E. VERNON. London: Methuen. 1957. Pp. 216. 15s.

'Secondary School Selection' is the first of what, it is hoped, will be several volumes on topics of public interest, published under the auspices of the British Psychological Society. It is the report of a working party of fourteen psychologists set up by the Society to give an authoritative statement of psychological research and opinion on the subject. The authors discuss at length the psychological, statistical and technical aspects of selection, but aim nevertheless to show that its problems arise primarily from historical, administrative, political and social causes, which cannot be removed simply by changing or abolishing existing methods. The report is addressed to educational administrators and teachers as well as to psychologists, and ends with thirty-two recommendations or spurs to action which serve to remind us that contemporary psychology is a profession as well as a field of academic study. It is chock-a-block with facts, arguments and theories, in which the influence of Prof. Vernon, who has edited the text, is clearly to be seen. Few psychologists will agree with all the findings put forward, but no one familiar with the problem could fail to appreciate the thoroughness and objectivity with which this task has been carried out, and it has certainly succeeded in showing how large a contribution psychology has made to this problem of national importance.

This book should go a long way to dispel the notion that the eleven plus examination is strongly supported by all psychologists, most of whom, it asserts, dislike the system, and look to alternative types of school organization based on the concepts of allocation (blessed word) and guidance rather than competitive selection: psychologists would apparently abolish the selection system if there were a practicable and just alternative, and many are said to have gone so far (and more haven't) as to advocate that their profession should 'contract out' and refuse to have anything to do with procedures that have such harmful effect on children's emotional and educational development. The committee takes the view that it is for society and its educational leaders to decide *whether* modifications in the selection procedure are desirable. As social scientists, psychologists should not, it is thought, feel committed to attack or to support the present system. Their function is to investigate which features are psychologically beneficial or harmful to individuals, and to ensure that the procedures employed are the most appropriate and valid for a given set-up. This gives a clear opinion on the vital problem of the social responsibility of the educational psychologist, who should presumably proffer advice to education authorities but should not thereby regard himself as any more privileged to insist on its implementation than the man in the next polling booth. This seems wise in view of the social consequences attendant on grammar school entrance, but there must surely be limits to the extent to which some psychologists will wish to co-operate in selection procedures which they believe to have had psychological effects on the candidates. It is only too evident to psychologists who come into contact with parents in clinics, or in lecturing to parent-teacher associations, that the natural desire of parents to ensure the best opportunities for their children has helped to make the process of selection for secondary schools a source of widespread anxiety. The report errs perhaps in giving a rather piecemeal emphasis to clinical experience concerning the effect of the examination on the four-fifths of the population who fail it, even if the emotional effects on the children are probably less than the ill-feeling caused among the parents. It might well be worth exploring such propaganda techniques as the film and pamphlets, which would help to bring home to parents what is involved in educational guidance at all stages of school life, and a shortened version of the present report, written for the ordinary parent, might perform an even greater public service than the present document. The working party thinks that, in view of the disquiet whipped up by the press and by politicians, exaggerated statements should be corrected. In this respect, many psychologists prefer to ignore ignorant criticism and to believe that the slow accumulation of objective evidence of the sort presented in this report will eventually reach the public.

Educational administrators will find advice on numerous topics, including scaling techniques, the use of subjective data, age allowances, the expected number of grammar school misfits, the marking of essays, school records, teachers' estimates, the selection of children attending small schools, the delimitation of the border zone (50th to the 95th percentiles), various borderline procedures, individual and panel interviews, an account of the extremely thorough method used in the city of Lincoln, and the West Riding scheme for scaling against the junior schools' results in previous years, or retrospectively against the

pupils' performance on tests given when they reach the secondary schools. The appendices might reasonably have included some material from the main text, as the discussion of such technicalities as multiple criteria and correction for attenuation and homogeneity are of little general interest, although everyone would be well advised to look at the appendices, if only to gain some insight into the complexities involved in selection. For example, Appendix B, on the constancy of the I.Q., develops the argument that the amount of variation will depend on the similarity of the tests employed, the time interval between testing and retesting, the chronological age of the subjects, the range of validity in the group, the level of ability, the short term reliability of the tests concerned, the standard deviations of the tests, the number of tests compared, and the range of the norms. The report emphasizes the very convincing evidence found in nearly every investigation of the superiority of the intelligence test as an overall predictor of later success, but fails to bring out clearly enough the additional advantage that fluctuations in I.Q., although larger than originally suspected, will usually be less than those in Arithmetic, English or any other selection variable. The important point is made that intelligence tests act much more as a class leveller than as a class perpetuator and that it would be more logical for left wing critics to welcome them than to decry them, since on both theoretical and experimental grounds it seems likely that in areas where the intelligence test has been abandoned fewer working-class children will gain grammar school places than in those areas where they still constitute part of the examination. Authorities who abolish intelligence tests are probably making a decision to withhold educational opportunity from working-class children.

Teachers will be interested to know that, in an important investigation, the scaled teachers' estimates gave correlations of the same order as those of the external examinations, and higher than those of the objective attainment tests, although the likely error of a teacher's unscaled estimate of the average ability of his class is twice that of the error of an estimate based on a mediocre intelligence test. 'Old type' examinations, when constructed by persons with experience of psychometric technique, can reach the same efficiency as the objective attainments test; and the essay, despite the unreliability inherent in its marking, does add appreciably to the validity of prediction, although no weight should be attached to a single marking of a single composition. The advice given about coaching is rather ambiguous. It is considered that coaching should be discouraged, but that if it cannot be controlled, adequate previous practice and explanation in all schools should be authorized. It is also admitted that the abilities of individual children are more variable than was earlier supposed, and that rigid streaming may have harmful educational and social consequences, although these are often exaggerated. And grammar school masters will, no doubt, agree that the G.C.E. is the nationally accepted standard for entry into higher forms, industry, commerce and the professional world, and so constitutes an important criterion of achievement.

The report will also serve as a text-book for students, and as a guide to lecturers in educational psychology. It includes sections dealing with the history of intelligence testing, theories of the nature of intelligence, the work of Hebb, the effects of environment and heredity on test scores, and the normality of I.Q. distributions. The excellent suggestion is made that the term 'general educability' should be substituted for 'general intelligence', on the ground that this term makes no claim that psychologists are measuring purely innate ability. On the other hand, it would be even more misleading to claim that we are measuring purely acquired attainment. And the work of Newman, Freeman and Holzinger is quoted merely to show that in very different environments twins with the same hereditary equipment might differ by as much as 20 points in I.Q., despite the fact that the main conclusion reached by these authors was that about 75% of the variance shown by tests of intelligence is attributable to innate or inheritable causes. Nor does this section on 'How Far Do Intelligence Tests Measure Inborn Ability?' refer to the extensive work of Burt in this field, which also points to the predominance of genetic factors. Research strongly suggests that in our present system differences in I.Q. scores are mainly determined by heredity. It is true, of course, that if our social environment were less homogeneous the environmental contribution would be increased, and that recent changes in the attitudes of parents, coaching, etc., may be altering the balance, but until contrary experimental results rather than hypotheses are advanced it would be more scientific to emphasize the innate component in intelligence test scores. Everyone would agree with the working party that psychologists should frankly acknowledge that a completely accurate classification is impossible at eleven years, but they should also add that perfect prediction of complex human behaviour at any age is impossible, and that psychological principles can no more be applied unerringly to individual children than the principles of physics can afford predictions of the exact path of a slate that falls off a roof. The important point is made that the best vindication for a system of selection is that it admits of fewer errors than would occur with any other method, and that we are now near the

limits of predictive efficiency, which is a considerable technical achievement. Yet the claim on the dust cover that the procedure is accurate for some 90% of children may be disputed, since it assumes that the correlation of an imperfectly reliable selection procedure with a perfectly reliable scholastic criterion is close to plus 0.90. In fact, scholastic criteria are by no means perfectly reliable, and the effect of this correction for attenuation is to give a somewhat optimistic estimate of the efficiency of selection procedures, although the discrepancy is not very great. It must be remembered that (as stated in the text) a quarter of those selected, on present standards, do not deserve their place. And is it fair to state that in view of the high validities achieved we can hardly expect that 'tinkering with' selection procedure, e.g. including essays as well as subjective English tests or additional personality assessments, would produce any marked improvement? The addition of essays and personality estimates are scarcely 'tinkering', since they raise the reliability, mitigate backwash effects on the school curriculum, and may eventually replace the original tests without lowering the overall validity.

Statistical psychologists will enjoy the appendices. Appendix A, 'Follow Up Studies of the Validity of Selection', comprises tables from six representative researches. Appendix B, on the constancy of the I.Q., summarizes research results, and ends with a table showing the gains or losses in I.Q. points to be expected for identical tests over intervals of 1 month and 5 years respectively. Appendix C, on the reliability of scaling of primary school marks and estimates, criticizes some of McClelland's earlier conclusions, and Appendix D gives particulars of sigma and percentile scaling and discusses the methods used by various education authorities, including the very workable scheme adopted by the National Foundation for Educational Research.

A question raised in everyone's mind is whether this report is not bolting the stable door after the horse has gone, since secondary school selection seems likely to be dropped by several authorities. However, educational provision in this country is such that any immediate wholesale change seems improbable, and, in any case, the authors of this report believe that, if comprehensive schools become general, psychology will have at least as much to contribute to the techniques of diagnosis and guidance within such schools as it has to the more restricted aim of accurate selection.

F. W. WARBURTON

Theories of Personality. By CALVIN S. HALL and GARDNER LINDZEY. New York: Wiley. London: Chapman and Hall. 1957. Pp. xi+572. 52s.

Most teachers of personality theory would agree that there is no *one* useful survey of the existing important theories. This book is intended to make good the deficiency. In the author's words, 'It provides compact yet comprehensive summaries of the major contemporary theories of personality written at a level of difficulty that is appropriate for undergraduate or graduate instruction.' It is their hope that 'this volume will serve a function in the area of personality similar to that served by Hilgard's *Theories of Learning* in the area of learning'. Chapters of a modal length of 40 to 50 pages are devoted to the theoretical positions of the following: Freud; Jung; 'Social Psychological Theories' (Adler, Fromm, Horney, and Sullivan); Murray; Lewin; Allport; 'Organismic Theory' (Goldstein, Angyal, Maslow, and Lecky); Sheldon; 'Factor Theories' (Eysenck, and Cattell); 'Stimulus-Response Theory' (Dollard and Miller, Sears, and Mowrer); Rogers; and Murphy. There is also an introductory chapter on 'The Nature of Personality Theory' and a concluding chapter on 'Personality Theory in Perspective'. A praiseworthy attempt has been made to outline each theory in terms of a standard pattern of exposition: introduction and orientation (some of the biographical material here is most interesting); the structure of personality; dynamics of personality; development of personality; characteristic research and research methods; and finally, current status and evaluation. But this expository schema has not been rigidly adhered to when not suitable, e.g. with Sheldon's theory. Also, when a number of different writers are taken as representing one theoretical position, as are Adler, Fromm, Horney, and Sullivan, there is only one section on 'current status and evaluation'. This is sometimes to be regretted, as when it involves no separate evaluation on Eysenck's position. Where possible, each theorist has read the chapter devoted to his theory; this includes nine out of twelve chapters. To this the authors acknowledge their heavy indebtedness. Furthermore, parts of the book have been read and criticized by an impressive array of big names in personality theory, and the whole book has been seen by Allport.

This book will certainly prove invaluable to both teachers and students of personality. Here at last is a single book which expounds most of the important contemporary approaches to personality in a mature and reliable way. It should prove especially useful as a basic text for honours students doing courses in personality theory; in this respect at least, its avowed aim of becoming the Hilgard of personality theory

does not seem immodest. At the same time its extensive references to both primary and secondary sources for each theory should enable the student to find his way into the literature. Criticism seems likely on two main counts. First, it is possible to disagree with the author's selection of theories; it might well be argued, for example, that the inclusion of a chapter on 'Organismic Theory' at the price of omitting such an historically important position as McDougall's (or indeed Shand's) and of devoting less than three pages to Mowrer is distinctly odd. There are those, too, who would question the wisdom of extended expositions of the views of Rogers and Murphy. But probably no two psychologists would agree completely about what to include and exclude. Secondly, it may be claimed that the authors' biases are from time to time too clearly evident, as for example in their somewhat lyrical account and appraisal of Murray's theory. On the whole, however, their expositions of the various theories are lucid and balanced; the chapters on Sheldon and on Lewin struck the reviewer as being especially good. The authors do not claim to present a full critique of each theory: their aim is essentially expository. Nevertheless, their sections on 'current status and evaluation' at the end of each chapter, though brief, are sometimes provocative; an instance is provided by their chapter on 'Factor Theories'.

Perhaps, the most remarkable and unusual quality of this book is the relative sophistication of its approach to the functions of a theory. This is especially evident in the first and final chapters, but also from time to time in the appraisals of the various theories (though here less often than one might wish). It is the authors' opinion that 'the field of personality would benefit enormously from an increased sophistication on the part of psychologists concerning the nature and function of theoretical formulation'. In the first chapter they outline a view of theory that is 'a relatively conventional summary of the thinking of methodologists or logicians of science'. Thus, a theory is looked upon as a set of conventions rather than as a fact of nature: the role of assumptions and of empirical (co-ordinating or operational) definitions is stressed, and the predictive functions of a theory are heavily emphasized. In the body of the text a constant criticism of theories is that they too often provide after-the-fact explanations and too seldom before-the-fact predictions. But the authors are not over preoccupied with *formal* considerations of theory adequacy, though they do emphasize that no existing theory measures up to these exacting standards. They draw a useful distinction between the way in which a theory may generate research by yielding, deductively, testable propositions about stable relationships between observables (classical hypothetico-deductive method), and the way in which, contingently, a theory may have heuristic value. For example, the authors believe that Lewin's theory, though formally inadequate, was heuristically very valuable, because it happened to inspire much research of generally accepted worth.

Another valuable feature of the book is the attempt made to compare the different theories with one another in terms of their relative positions concerning a dozen and a half important points of controversy on the content (rather than the form) of personality theory. Examples include the varying importance of purpose, of unconscious motivation, of reward, of contiguity, of heredity, and of the concept of the self, in the various theories. A table summarizing this 'Dimensional comparison of Theories of Personality' is given on p. 548. Interesting, too, as well as provocative are the authors' 'Some Reflections on Current Personality Theory' in the last chapter. What is particularly commendable here is the way they stress the value of empirical research that bears directly on existing theories: they castigate the current over-concern with theoretical originality, however trivial. Their rejection of the attractive, but facile, view that present theories of personality should be 'synthesized' is also refreshing.

In all, the authors are to be congratulated on producing a sophisticated, relatively well-balanced and reliable survey of a difficult, diffuse and highly controversial field of psychology. They will certainly receive the gratitude of many future students and not a few teachers.

C. W. PILKINGTON

Perspectives in Personality Theory. Ed. by HENRY P. DAVID and HELMUT VON BRACKEN.
London: Tavistock. Pp. xiii + 435. 38s.

We are told in the last chapter that 'this book is a child of the Fourteenth International Congress of Psychology, which convened in Montreal in June 1954'. The 'child' proves to be a wayward adolescent, full of conflicts, with an unintegrated personality—now sanguine, now aggressive, now confused. In so far as he has a unity, this derives from his preoccupation with national and 'school' differences in outlook, on the question of personality theory.

The book consists of 22 chapters each by a different writer. Eleven are listed in the Introductory Note as psychologists from the United States, the other half being drawn from Belgium, Canada, England, France, Germany, Holland, Italy and Switzerland. Despite the evident efforts of the contributors and the

editors towards topicality, the reader increasingly gains the impression that *plus ça change plus c'est la même chose*. The changing and occasional coining of 'technical' terms merely emphasize the inevitable recurrence of the concepts. There is much play with the traditional antithesis between 'philosophical' (or speculative, deep, characterological, understanding—or European, i.e. German) on the one hand and 'scientific' (or experimental, objective, behaviouristic, explanatory—or American) on the other. One of the useful functions of the book, however, is to demonstrate implicitly how untenable, as well as sterile, such dichotomies are. First, it is obvious from contributions on both sides of this straw fence that clarity and experimentation are not the prerogative of the 'scientific' psychologist, nor obscurity and special pleading the hall-mark exclusively of the philosophically-minded. Secondly, a glance at the names and present home of the contributors points the absurdity of any classification nowadays based on geographical or national considerations. The United States are represented, for instance, by Else Frenkel-Brunswik and Werner Wolff just as they are by Gordon Allport and Frederick Wyatt; the two U.K. contributors are Hans Eysenck and Cyril Franks; Otto Roser hails from Italy, Wolfgang Luthe from Canada.

The book is divided into six sections entitled respectively: Overview, European Trends, Theory, Methodology, Commentary and Resources. The Overview consists of Allport's chapter. This is mainly a discussion of what he calls the 'Anglo-American' and 'Continental' approaches. 'Resources' comprises a useful, well-classified Bibliography (324 references) by David and a Note on the International Congress by MacLeod.

Of the four central sections which form the body of the book, the reviewer found most stimulating the chapters by van Lennep, 'Projection and Personality', Wolff, 'New Concepts in Experimental Depth Psychology', and Frenkel-Brunswik, 'Perceptives in Psychoanalytic Theory'. These three illustrate the practicability of combining 'understanding' with 'explaining' and even 'predicting'. There are also two interesting chapters in the Commentary by, respectively, Wyatt and McClelland. In spite of the bird's-eye view which these two writers attempt, they manage to be both constructive and tolerant.

A. W. HEIM

Psychology in the Soviet Union. Edited with an Introduction by BRIAN SIMON. London: Routledge and Kegan Paul. 1957. Pp. viii + 305. 32s.

Two years ago, after a visit to the Soviet Union which included discussions with several Soviet psychologists, Mr Simon decided that there was need for a book in English containing samples of recent Soviet work. He put this suggestion to Prof. Smirnov, the Director, and Prof. Menchinskaya, the Deputy Director, of the Institute of Psychology in Moscow, and they immediately responded by sending him thirty papers recently published by Soviet psychologists. The present volume contains edited translations of twenty-two of these papers, together with a concise and illuminating survey of the whole field by Mr Simon. None of the papers is more than 6 years old, and all but four were first published during the last 3 years.

The papers are arranged in five main groups, according as they have been judged to be primarily concerned with: the general direction of research in Soviet psychology; Pavlovian findings and principles; perception and memory; learning and development; or general psychological theory.

While the papers are valuable for their detailed discussions of particular subjects, they also provide a general picture of current Soviet psychology. Some of its main features may be briefly indicated. First, orthodox Soviet work is firmly based on the principle that mental processes and behaviour must be explained by reference to the neural processes that underlie them. This is 'the only correct view' (p. 29), and although some recent articles have deviated from it, they 'aroused keen opposition from the general body of Soviet psychologists' (*ibid*). Secondly, the development of an individual is assumed to be determined by his 'conditions of life' and his 'activity'. The influence on development of the 'conditions of life' must not be understood in 'an over-simplified mechanical way' (this is 'the error of the behaviourists'); it involves 'the whole complex dynamics of the cortical processes' (p. 47). Thirdly, all consciousness is held to be a reflexion of the objective world, including, in Pavlov's words, 'the entire organism and all its component parts', and it arises as the result of processes of analysis and synthesis at various levels in the nervous system, and especially in the cerebral cortex. (But, strangely enough, there is no reference in these papers to information theory—or, for that matter, to cybernetics.) Fourthly, Pavlov's theories, and even his *obiter dicta*, are usually accepted as authoritative, and, in the explanation of various psychological processes, there is considerable uncritical use not only of conditioned reflexes but also of the 'second signal system' and other mechanisms involved in his account of 'higher nervous activity'. Fifthly, mental tests and all other psychometric techniques for assessing individuals are roundly

condemned. It is charged against them that they presuppose fixed differences between individuals, that they are unreliable, and that they are 'restricted to a mere statement of the result, without disclosing why or how it was achieved, or what prevented it from being better' (p. 189). Sixthly, except in their reports of international congresses, Soviet psychologists rarely refer to 'the psychology of the West', and, when they do refer to it, it is usually to condemn psychoanalysis and Gestalt psychology for their subjectivism, behaviourism for its over-simplified mechanist standpoint, neobehaviourism for its neglect of psychoneurology (and its undifferentiated use of 'habit'), and 'so-called social psychology' for its 'superficial inquiries', which have 'no scientific significance whatever' (p. 295). Finally, Soviet psychologists, like Russians in other fields, practise severe self-criticism—self-criticism, that is, in terms of the orthodoxy to which they are expected to conform. Thus Prof. Smirnov summarizes (pp. 41–5) nine 'important gaps and deficiencies' in Soviet research. He says, among other things, that 'theoretical work is inadequate', that 'there is an almost complete lack of studies of the psychology of labour', that 'there has been a significant reduction in work on psychopathology', and (as we have heard of psychology nearer home) that 'the scientific level of research is not always satisfactory: some investigations have no relation to either theoretical or practical tasks; others are superfluous; certain conclusions sound like well-known truisms'.

Mr Simon tells us that in the Soviet Union psychology is taught to all pupils during the final year of their 10-year schooling, that it is available to undergraduates in 'the relevant departments' of the universities, that it forms part of the course for all intending teachers in universities and training colleges, and that it is developed both in special research institutes of psychology in Moscow, Leningrad, Kiev and Georgia (presumably in Tbilisi) and in research institutes in education in all the republics. The papers in this book all come from Moscow, Leningrad or Kiev, but they sometimes refer to psychologists in other parts of the Soviet Union (though not to the Georgian followers of Uznadze, who are said by *Soviet Culture* to be involved in heterodox excursions into 'Freudism'), and we may infer that, although most Soviet psychology is to be found in these three cities, there is also something, but not perhaps very much, going on elsewhere.

In all the papers there is evidence of the influence of the ideological strait-jacket. Everything is made to conform strictly to officially approved Pavlovian principles; there is explicit submission to 'the Decree [regarding the influence of heredity and environment on the development of an individual] of the Central Committee of the Communist Party of July 4, 1936' (p. 47); there is repeated reference to the sacred texts—the 'classics of Marxism-Leninism' (p. 133)—and to such particular Marxist doctrines as that 'sensations, which are inseparable from man, developed historically in the course of development of social labour' (*ibid.*); and, in a paper first published in 1954, there is solemn acceptance of Stalin's pronouncements on the nature of language and its relation to thought. Again, while valuable advances are being made in psychoneurology, which Soviet psychologists sometimes call 'the neurodynamic explanation of mental processes', much of the research in educational and child psychology (and there is apparently more research in these fields than in any other) is concerned with specific, practical problems rather than general, basic issues. Still, even here, Soviet psychology has important things to say about the role of learning, and of social factors, in perception, motivation and the development of personality; and, as these matters have only recently become prominent in British and American psychology, it is useful to see what is being made of them in the Soviet Union, where the ideological climate has long led to their being among the principal subjects of study.

Clearly, this is an informative and valuable book, even though the value of the collected papers perhaps lies less in their contribution to psychological knowledge than in the picture they present of the main characteristics and tendencies of Soviet psychology as it was 2 or 3 years ago.

REX KNIGHT

Psychological Tests and Personnel Decisions. By L. J. CRONBACH and G. C. GLESER.
Urbana: University of Illinois Press. 1957. Pp. xii+165. \$3.50.

Using the concepts and techniques of Decision Theory the authors tackle the problem of selection and placement with the most interesting results. The term selection covers not only ability tests but also interviews and personality tests; placement refers to educational, industrial, and psychiatric placement. The analytic device used in this application of Decision Theory is the comparison of strategies, i.e. the policies used to select and place people. The object is to choose the strategy that is most beneficial to the institution concerned. By using pay-off functions the authors show how it is possible to compare the utility of training programmes over the whole range of a given ability when the conventional methods only give

the best programme without any information on the usefulness of the programmes for different levels of ability. For developing training schemes such detailed knowledge would be invaluable in making policy decisions on the relative advantages of grouping people in terms of ability. Perhaps the most important contribution this new approach makes is to query the adequacy of the present methods of validating tests. The coefficient of determination is found to apply only under certain conditions and the coefficient of forecasting efficiency is suggested as having no relevance whatsoever. The present methods of validating tests do not take into account what the authors call the *a priori* strategy, i.e. knowledge such as academic records which could be used as a means for selection and placement without reverting to tests. They maintain that it is in relation to this that a test should be validated. Furthermore, before deciding to use a test the advantages which would accrue from adjustments to training programmes and job simplification must be considered since they are shown to be intimately associated with the value of the validity coefficient. How much a test should be lengthened to increase its validity must be balanced with the increased cost of the test; equations are given for finding the optimum length of a test for this condition. The authors conclude there is no single validity coefficient; a test must be considered in relation to the objectives it has to attain. Selection using single and multiple tests is analysed to provide equations for determining the optimum cut-off scores and selection ratios. Another important question discussed is whether to measure a small number of aptitudes efficiently or a large number with moderate efficiency. The problem as seen by the authors is a matter of balancing the two extremes to obtain maximum utility from the tests. Again the problem of validity is encountered and it is pointed out that the contribution to utility of a number of moderately efficient tests may be greater than a smaller number of highly valid tests. Classification problems are taken as the general case of selection and placement but the topic is not developed because it entails numerous restrictions. To evaluate the outcome of a strategy requires partly or wholly subjective judgments, and is, therefore, open to criticism. This weakness is no greater than that of other psychological theories; on the contrary the explicit statement of outcomes makes it stronger than most.

This stimulating book has brought much needed cohesion to the subject and has overcome some of the problems that have faced conventional test theory. Whether the solutions it advocates will be accepted by its readers will depend not upon the rigour of the mathematics, which cannot be faulted, but upon the assumptions from which the mathematical arguments are developed. The generality of the new approach will appeal to the mathematically inclined psychologist but the consequences of it are important to all applied psychologists. The empirical work entailed in the development of tests in accordance with the theory will be more exacting but at the same time more rewarding than that for conventional theory. Decision Theory is potentially a powerful tool for exploring new areas of psychology. To apply it, however, requires mathematical competence and a sound knowledge of the subject-matter to be explored; both qualities are shown by the authors.

For applied psychologists this book is an absolute must because it deals with classification problems pertinent to their work. It can be recommended to all psychologists as being amongst the studies extending psychological theory into new and important fields. Whilst at times non-mathematical readers will find the arguments a little difficult to follow they need not be abashed by the book since the authors put the minimum of formulae in the text, preferring to give the mathematical development of the theory in appendices. Before reading this book most readers will find it helpful to read *Design for Decision* by H. D. J. Bross (New York: Macmillan Co., 1953) which gives a good popular introduction to Decision Theory. There are two misprints in Appendix 3, p. 142; they concern formulae (3.3) and (3.4).

H. C. BAKER

Information Theory: Third London Symposium. Ed. by COLIN CHERRY. London: Butterworth. 1956. Pp. xii+401. 70s.

A book of 400 pages presenting papers on forty different topics confronts the reviewer with something of a problem. Information Theory is not a tidy and coherent body of systematically related concepts, nor is it a great inductive generalization from convergent experimental studies. Or at any rate if it is either of these things the fact does not emerge from this symposium. A hostile reviewer would accordingly have an easy task, by selective quotation, in making mincemeat of the whole project. But he might thereby reveal more of his own limitations than of the significance of the work in question. It is probably not unfair to say that Information Theory represents a movement of youngish men, even though the pioneers are now venerable. There is vigour, there is enthusiasm, there is scant regard for the established inter-

subject boundaries of the past, there is preference for the fertile suggestiveness rather than formal conclusiveness, there is reliance on teamwork and mutual intellectual support—the characteristics, in fact, of Operational Research. The social psychologist and the anthropologist will discern a number of familiar tendencies in such a movement—it has features in common with a nascent religion.

The hardest aspect of the movement for those reared in a slightly earlier discipline is not its mathematical dressing nor even its conceptual complexity but the ambiguity which saturates every discussion. If you have regularly had your knuckles rapped for loose language, whether in logic or in linguistics or in methodology or in philosophy, it is hard to adapt yourself to an atmosphere in which quite other standards prevail. It is like being brought up in a society adhering to strict matrimonial orthodoxy and moving into one where free love is accepted. You may try to adapt, you may go through the motions of a new conformity, but your conscience makes you miserable. And if you have a sneaking belief that the new society is, in important respects, better than the old your distress is all the more acute. These information-theorists may be swimming in a sea of ambiguities but they seem to be getting somewhere.

It is important not only to assess what Information Theory is saying but, even more important perhaps, to appreciate the innovation in methodology implicit in the fact that so many very different things are being said in what purports to be a single context. It is a context in which speech jostles with thermodynamics, chance with electronics, librarianship with arithmetic, games with linguistics, taxonomy with optics, perception with induction, logic with neurology—in fact almost anything goes. And out of this higgledy-piggledy empire we are getting, among other things, automation. Its own conceptual unity, and, in time one hopes, its own coherent discipline must take shape out of its seemingly amorphous assemblage of diverse elements.

Information Theory is already making piercing inroads into most of the existing major disciplines—into logic and mathematics, physics and biology, linguistics and psychology, medicine and technology. Let us not underestimate the magnitude of its impact. It is unlikely that anyone can as yet correctly interpret its significance. It is akin to one of these shifts of ecological equilibrium in which the interacting factors are so numerous, and so resistant to precise calculation, that the outcome is quite unforeseeable. We will each look for what interests us most in the phenomenon. The practical effects are more immediate and more obvious than the long-term theoretical implications. One of these latter is surely the methodological discovery that ambiguity is not only inevitable in all language but that it may be a principal source of novel and usable concepts—just as free love might break up the genetic stratification of a rigid social system in which caste separates marriage-groups as effectively as geographical distance. Novel genetic clusters will emerge, and similarly Information Theory is throwing up quite novel conceptual clusters.

The importance of ambiguity lies in the light that it throws on interpretation. We have been obsessed too long with an ideal of one-one correspondence between elements in the stimulus and elements in the percept. The realization that stimulus-elements serve as cues for the evocation of interpretative framework is not novel but its implications go much further than the theory of perception has indicated. The psychology of productive thinking is involved here, and with it the psychology of communication. The terminology of 'signal' and 'noise' borrowed from telephone engineering is useful here and has been put to good effect in this symposium by D. J. Bruce of Reading in a simple experiment capable of considerable development. Four series of tape-recorded words are presented a number of times to a group of subjects. They are combined with noise at varying levels, greatest at the first presentation and progressively diminished, with of course a resulting increase in word-recognition. Two of the word-series were of a random character. Each of the other two was selected from a specific context namely, *parts of the body* and *kinds of food*. The outcome, in terms of graphs of recognition-performance, is, of course, what one would expect. The interest lies in the detailed light thrown on the recognition-process. We tend to speak of 'context' as if it were something appertaining to words. When we appreciate the fact that the context must inevitably be supplied by the subject we see the storage-system of the brain in a new light. It not only stores 'bits of information' but a hierarchy of established contexts. And it must possess mechanisms for switching freely into one context after another in its groping after interpretations. It is perhaps unfair to select for mention a single paper out of a symposium of forty papers but there is not only the excuse of psychological interest and of limitations of space but the further justification that this experiment, rightly interpreted, will be found to exemplify most of the major concepts occurring in the Theory of Information.

G. PATRICK MEREDITH

n Human Communication. By COLIN CHERRY. New York: Wiley. London: Chapman and Hall. 1957. Pp. xiv + 333. 54s.

This book is intended as a general introduction to a series of forthcoming texts by various authors on specific aspects of communication. It therefore touches on each of the sciences connected with the subject, in a series of self-contained chapters: one introductory, one historical, and the remainder divided between linguistics, phonetics, statistical communication theory, semantics, and psychology. Each chapter is meant for readers who are not specialists in the particular topic, so that the book may well attract those who want an outline of the ideas of neighbouring disciplines. A number of references are given, which allow further exploration: their value is reduced by the absence of an author index or any other way of finding the portion of text in which a particular reference is cited.

It would be easy for any professional psychologist to find fault with the chapter on that subject: certainly those of, say, the learning theory variety should be warned against regarding the last chapter as a comprehensive guide to the activities of their cybernetic colleagues. But such criticism misses the point of the book. We ought rather to ask whether it succeeds in communicating ideas across the barriers between the sciences. The strongest parts, as might be expected, are those dealing with communication theory and signal analysis. On these topics more is expected of the reader than in other chapters: he can hardly survive if he is frightened of an integral sign, and many psychologists are. But easier short treatments are already available; the merit of this one is rather that it is intermediate, going into some detail without the length and complexity appropriate to an engineering handbook. The other disciplines are handled adequately on a more elementary level.

All in all, this is probably the best introduction to all these neighbouring fields yet available in a single volume. There are few people who could have written it single-handed. However, Dr Cherry's broad sympathies are only too clearly revealed by his remark that 'the mechanical engineer designs a motor-car to operate with its driver, as an integrated man-machine unit'. Too often the engineer does nothing of the sort, and the car-driver suffers. Equally many psychologists are content to regard language as merely a kind of bee-dance. In this case it is psychological theory which suffers. Books of this sort serve to remind us that there is more behaviour in our daily life than is dreamed of in our philosophy.

D. E. BROADBENT

Interviewing for the Selection of Staff. By E. ANSTEY and E. O. MERCER. London: Allen and Unwin. 1956. Pp. xiv + 111. 10s. 6d.

Two experts have collaborated to produce one of the few really effective aids to those who, often without qualifications and training, have to do a job in the field of applied psychology. It is a popular account of what the layman will find it easy to read but it is certainly not popular in the pejorative sense of that word. There is nothing in it that a university psychologist would not be proud to have written.

After modestly limiting the purpose of the book to those who are occasionally called on to interview people for appointments or promotions, a few pages are devoted to the aim of the selection interview. The next chapter advises the interviewer on how to ask himself questions before interviewing, questions about aspects of the job and then questions about the candidates to which he might hope to discover answers by means of the interview. A short chapter on 'Preliminaries and Setting' bears the stamp of the authors' experience. A miracle of condensation is achieved in the heart of the book in a 10-page chapter called 'Method during Interview'. It could not be bettered in such a small space.

Excellent advice is offered on note taking and report writing, and specimens of rating scales and interview forms are shown. Critical and knowledgeable accounts are given of Board Interviews and Multiple Interviews and good sense is talked about hauling all the material together into conclusions and recommendations. A novel feature is a chapter entitled 'Advice to a person about to be interviewed'—and good advice it all is. Here is a sample. 'You may be asked about your family. Whatever your origin, refer to it openly and without defensiveness. If your father had a distinguished career, why conceal the fact? If he was poor or for a time unemployed, you may be the more able to pay tribute to the way your parents brought you up.' For good measure three appendices and a booklist are added. A bargain at 10s. 6d.

DENIS MCMAHON

Supervisory and Executive Development: A Manual for Role Playing. By N. R. F. MAIER, A. R. SOLEM and A. A. MAIER. New York: Wiley. London: Chapman and Hall. 1957. Pp. xi+330. 52s.

For those who wish to indulge in role playing as part of human relations training, this book provides complete instructions. Its authors claim that it is equally applicable to industrial executives and supervisors, individually or in small or large groups, and to the family circle. No skilled instructor is required for it does not set out to be a text-book; it 'proposes to avoid giving answers. Some principles and facts may be introduced, but basically the aid that may be supplied is in the form of raising questions and exploring issues'. The course of twenty case-studies covers the ground from straightforward, everyday problems met by first-line supervisors to complex organizational questions confronting managers and directors. Eight of these studies appear to consist of new material, the remaining twelve being adaptations from work published previously by the senior author. Only one case requires more than eight players, and six of them contain only two parts. However, minutely detailed directions show how groups of the order of a dozen up to several hundred members may derive simultaneous benefit from the programme. For each case there is an introductory statement focusing the problem to be solved. Next come instructions for the conduct of the 'play' (either Multiple, or Single Group, Role Playing Procedure, including sometimes the Skit Completion Method or the Dramatized Case Method), followed by background to the specific study and roles for the players. Finally, there is a discussion about the general human relations principles involved.

It will be clear from this account that the book is likely to be of value to serious students of psychology and sociology only in so far as they may wish to familiarize themselves with these particular applications of role-playing techniques; it is not written for them. Its potential field of usefulness includes, rather, the various extra-mural, W.E.A., technical college and adult education centre courses concerned with human relations, and training departments in industrial organizations. Except possibly for the last of these, the price will be a serious drawback, since each member of the group requires a copy; the roles are deliberately not available separately, and reproduction is expressly forbidden. For groups of all these kinds, some rewriting would be needed to make the settings more realistic, and the names and vocabulary more familiar (the word *unique*, for example, occurs at least four times in the first 10 pages). In fact, the book may well be of more value as a starting point for an instructor than as a manual for class members.

On the one hand we may feel grateful to the authors for having given so careful a picture of one aspect of contemporary American supervisory training practices; on the other, we are left with regrets that no evidence is included as to the effectiveness of role-playing techniques in furthering one of the most important aims of such training—the change of attitudes in actual work situations. Nor is the need for such evidence mentioned.

DERMOT STRAKER

Psychology of Personnel in Business and Industry. By ROGER M. BELLOW. London: Staples. 1956. Pp. xii+467. 42s.

This thick book is the first British publication of the second American edition. It is written more for business executives and personnel managers than for psychologists but although briskly written is not at all a popular handbook. Part one is a very interesting but, to British readers, not very useful layout and description of the kinds of work different kinds of personnel people do in U.S.A. Part Two, about one-third of the book, reports research on workers' satisfactions, group dynamics and attitude surveys and covers in a condensed but effective way the material used by Viteles in his *Motivation and Morale in Industry*. This Part has grown and has been brought up to date since the first edition. Those whose acquaintance with the literature on the theory of social and industrial psychology stops at Roethlisberger, Lewin and Moreno will find this part of the book very useful.

Part Three is an account of modern selection methods and of some aspects of training. In this Part the most interesting chapters for this reviewer were one on Recruiting, which sets the whole problem of selection in a wide context, and another on Readability of training material in which much more than Flesch's work is reported. (I particularly liked the formula for calculating the 'Fog Index' of a piece of writing.) After finishing the book one has a fairly clear picture of the author: he seems to be an efficient, keen, executive-type industrial psychologist who is research minded and a scientist at heart. The book pretends not to be scholarly (presumably so as not to put off the kind of reader for whom it is written) but

is packed with research material and does not stray beyond the evidence; one often discovers that the author has himself carried out applied research in a particular field although some of it has not been published. Psychologists working in the social-industrial fields will find it a useful book, students interested in psychology applied to industry should have access to it and all British personnel managers would find it very valuable.

DENIS MCMAHON

Intelligence in the United States. By JOHN B. MINER. New York: Springer. 1957. Pp. xii + 180. \$4.25.

This book is a strange mixture of over- and under-statements. The author claims that he has made a survey of a cross-section of the U.S. population 'with conclusions for manpower utilization in education and employment'. This would appear to be overstating the case since the book is based simply on the results of testing 1500 subjects (1347 whites and 153 negroes, ranging in age from 10 years to 75+) on the 20-word, multiple-choice vocabulary test, and comparing their scores with their occupations ($n=316$), race ($n=2$), religion ($n=4$), social class as self-estimated ($n=4$) and sex ($n=2$). The other variables considered include education, place of residence, marital status, and age. On the basis of his subjects' scores on the vocabulary test and their position on these other variables, Dr Miner estimates what proportions of people in the U.S.A. are over- and under-employed, over- and under-educated. He expounds, in percentages correct to one—sometimes two—decimal place, how many should be 'promoted' and how many 'demoted'. It is among these conclusions especially that understating of a kind occurs. For example, having ascertained 'the degree of manpower wastage as indicated by tests of verbal learning' and specified the numbers at every level who should be promoted, he adds with apparent light-heartedness: 'In order to determine the extent to which our present occupational placement system produces a deviation from one based entirely on a verbal ability criterion, all that is necessary is to compare an individual's vocabulary score with the ranges that have been established and then determine whether it would be necessary to promote or demote him to ensure his employment at a level consistent with his intelligence. The degree of deviation can then be specified in terms of the percentage of the working population who would have to be promoted to a higher occupational level to attain a perfect intelligence-occupation correlation. An equal number will, of course, have to be demoted in order to keep the number of people working at each level constant' (reviewer's italics). The last sentence (which ends the section) strikes the reviewer as a masterly understatement of difficulties—leaving aside any question as to the defensibility and value of Dr Miner's plan.

Dr Miner is interested above all in avoiding wastage. Towards the beginning of his chapter on 'The utilization of intellectual resources in the educational system' he says: 'The data, to be presented in this and the following chapter, on educational and occupational manpower wastage are unique.' The reader, pursuing page after page of what can only be called autistic thinking, becomes inclined to agree. Dr Miner's conclusions rest on two assumptions. The first is the infallibility of the 20-word vocabulary test at all levels and for all purposes. The second is the fact—rather than assumption—that 90% of any given group score above the 10th percentile! Lest it be felt that this analysis is less than fair to Dr Miner, the following two exceptions are offered for consideration. (a) 'A score of seven comes directly above the tenth percentile point of the semi-skilled workers, and accordingly anyone with a score of seven may be considered as capable of semi-skilled work but not of skilled or highly skilled jobs. Anyone with six or less words correct is, by exclusion, intelligent enough only for unskilled labor.' (b) 'In establishing minimum requirements for performance at each educational level, the score directly above the 10th percentile was chosen... In using the 10th percentile point as a lower bound, a tendency to underestimate the number of students who could do more advanced work is introduced.'

The remaining parts of the book deal primarily with the findings of other workers who have investigated test scores as related to school, home and occupational environment; and with 'social stratification', the main variables being race (black and white) and social class. On the last two points, Dr Miner makes several naïve comments. For example, referring to social class he says: 'What leads people who do not meet the objective criteria for membership to identify with the upper class is impossible to say.' Towards the beginning of the book the reader has difficulty in deciding just where the author is going. By the end of the book, he has decided that it would have been better not to travel at all than to travel hopefully.

A. W. HEIM

Perceptual Processes and Mental Illness. By H. J. EYSENCK, G. W. GRANGER and J. C. BRENGELMANN. Maudsley Monographs No. 2. London: Chapman and Hall. 1957. Pp. 144. 25s.

The inquiry described in this monograph was undertaken in order to investigate the possibility that the perceptual processes of neurotic and psychotic individuals differ in some way from those of normals; to 'isolate differentiating variables and offer explanatory hypotheses concerning the processes responsible for differentiation'. Such an investigation could be of great interest and value, particularly if the nature of the differences could be determined. Unfortunately the experiments carried out by the authors of the monograph cannot advance our knowledge greatly. They presented a large and heterogeneous collection of tasks to groups of normal, neurotic and psychotic individuals. These included sensory tasks (dark adaptation, colour vision, acuity, after-images, etc.); oculo-motor tasks (binocular fusion, accommodation, etc.); perceptual tasks (Gottschaldt figures, visual illusions; number, pattern and object perception, etc.). In so far as can be judged from the brief and not very lucid descriptions given, the experiments were frequently carried out by techniques too crude to show up other than gross differences of function; they were not always adequately controlled; the readings were too few; and the scores were in many cases expressed in arbitrary units the physical values of which are not given. All we can conclude is that in many of these tasks the neurotics and psychotics performed slightly worse than did the normals, but that there was a great deal of overlap between the different groups for most of the tasks. An attempt was also made to show from some of the experimental results that a classification of the subjects in terms of Jaensch's 'integrate-disintegrate' typology gave some indication of a stronger tendency towards the 'disintegrate' type among abnormals than among normals. But the results were not very consistent, as is usually the case with perceptual typological classifications of this kind.

The scores of the normal group of subjects in the perceptual tasks and on certain non-perceptual tests (intelligence, motor, physiological) and questionnaires were correlated and submitted to factorial analysis. Five factors were extracted, the third of which had high loadings on visual acuity (as measured by the Snellen chart), fluctuation of after-images, perception of letters and of objects. In these the distribution of scores of normals and abnormals shows a comparatively small overlap, so that about three-quarters of the cases in the two groups were correctly differentiated. However, it is doubtful whether even these tasks could be used as diagnostic instruments as they stand.

A far higher degree of care and accuracy in the design and performance of the tasks, such as is normally used in carrying out experiments of this kind, is obviously necessary before any valid data can be obtained to demonstrate that neurotics and psychotics perceive differently in any consistent way from normals. Inter-correlation of the scores of the former in a number of different perceptual tasks might give some indication of this. But also the type of experimental task and the results obtained must be *psychologically* investigated and analysed before anything can be deduced as to the nature of the differences. Finally, it is probably important to take into account the effects of incentive conditions, since it is quite possible that abnormal subjects are poorly motivated to undertake tasks which are tedious to perform accurately and carefully, rather than inferior in perceptual capacities.

M. D. VERNON

Left handedness: Laterality Characteristics and their Educational Implications. By MARGARET M. CLARK. University of London Press. 1957. Pp. xvii + 214. 15s.

This monograph, sponsored by the Scottish Council for Research in Education, is divided into two parts, the first part reviewing the literature and the second reporting the results of an investigation into the hand, foot, eye and ear preferences of children drawn from eight Glasgow primary schools.

The literature on handedness and allied topics is extremely diverse, and the author has done well to provide such a concise and judicious summary of previous work. The explanation of hand dominance falls under two headings: granted a division of labour between the two hands, why has man become predominantly right-handed?; and granted this right-handedness, why does a select minority insist on deviating towards the opposite extreme? These two riddles remain unsolved, yet some ingenious theories have been put forward, most of which receive honourable mention here. A chapter is devoted to the genetic approach, and the development of laterality in young children is briefly outlined. The possible connexion between enforced change of handedness and stuttering, and between crossed laterality and reading disability is critically examined, while the special difficulties that are apt to beset the left-handed writer are given due prominence.

The author's sample of 330 children aged between 11-12 was unselected as regards laterality, and free from any known bias in respect of intelligence and social background apart from being entirely urban. A battery of 18 tests, most of them designed to measure preference rather than skill, was applied to each child individually, and a writing test was administered in the classroom. The results are analysed with painstaking attention to detail; and while it may not be easy for the general reader to keep his perspective as he gropes through a forest of statistics, the raw material is all there for anyone who wishes to discover how many left-footed hoppers were observed to kick with the right foot, or the ratio of left-eyed to right-eyed peepers.

Some of the conclusions drawn, e.g. that left-handers can hold their own both as writers and as scholars, will come as a relief to some parents, although probably as no surprise to workers in the educational field. When the author concludes finally that 'no essential difference was apparent between left-handers and right-handers, except for their use of a different hand' one wonders whether the matter ought really to have been allowed to rest there, for surely their behaviour in the test situation cannot be expected to tell the whole story. Indeed at least one reader, himself a member of that 'sinister minority' to which the author is evidently proud to belong, is doubtful whether any valid conclusions can be drawn about the significance of left-handedness from a sample which included no more than 18 left-handed writers. Yet while one might have wished for a more imaginative approach to the problem, a survey of this kind has nearly always something to contribute. All who are professionally concerned with primary education would do well to read it.

MICHAEL HUMPHREY

Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf. Ed. with an Introduction by JOHN B. CARROLL. London: Chapman. New York: Wiley. 1956. Pp. xi + 278. 56s.

Benjamin Lee Whorf was an industrial chemist, trained at the Massachusetts Institute of Technology, who attained a remarkable reputation as a 'trouble shooter' for a Fire Insurance Company. Early in his comparatively short life—he died at forty-four—he turned his unusual abilities to the study of linguistics, working for a while with Sapir at Yale. Apart from purely linguistic essays this collection of his writings contains a number of papers maintaining the thesis that 'thinking', in particular thinking about such ultimate metaphysical problems as the nature of space and time, derives from the language used by the thinker. The general thesis is not of course by any means new, but its derivation from American Indian languages such as Hopi will be interesting to psychologists. One reads the book with respect for the power and originality of the author's mind, but with the feeling that the psychological issues are treated from the outside, by one only superficially cognizant of psychological method and data.

GEORGE HUMPHREY

A New Approach to Figure Drawing. By L. CALIGOR. Springfield, Illinois: Thomas. Oxford: Blackwell. 1957. Pp. xii + 148. 34s. (Test materials available at 45s.)

Projective techniques utilizing human figure drawing are becoming more widely known and the present book adds another method to the approaches associated with the names of Goodenough, Wechsler and John Buck. The new technique is intriguing—the subject goes from one drawing to the next whilst the preceding drawing is visible to him through the transparent paper; he may trace it and virtually repeat it, modify it or ignore it completely. His handling of this situation, as well as the various details of the drawing, the placement and line quality are considered valid diagnostic clues.

Dr Caligor provides a scoring system based on objectively scorable dimensions and attains a high inter-scorer agreement. The scores are stated 'in terms of a deviation from a statistical norm approximately the mode scores of normal subjects'. Unfortunately the book contains no statistical material and the only relevant reference given by the author refers to a study of items differentiating between 21 college males and 21 hospitalized paranoid schizophrenics—this could scarcely be called a normative study serving as a basis for a complete test manual to be used in clinical practice.

Even more disappointing is the systematic list of suggested interpretations for the various scoring categories which are, no doubt, based on wide clinical experience, but which would be more convincing and valid if supported by some evidence. Though Dr Caligor's technique shares this fault with the systems devised by his predecessors, one feels that nowadays, having demonstrated the wealth of clinical

material obtainable from drawings, investigators should approach this field more systematically and not by-pass such important factors as sex, age, intelligence, cultural background, etc.

In the meantime, this new publication's main merit is the presentation of a novel and promising projective technique and of a scoring method which may well be used for testing experimentally the validity of drawing interpretations.

H. C. GUNZBURG

On the Early Development of Mind. By EDWARD GLOVER. London: Imago. 1956.

Pp. xi+483. 45s.

Dr Glover is a veritable Mrs Battle. He is all for a clear fire, a clean hearth, and the rigour of the Freudian game. He regards most of the neo-Freudian developments much as that good lady would have contemplated Bridge or Canasta, or any other sectarian variant of her favourite sport. 'The object of the following survey'—his 1949 paper on *The Position of Psycho-analysis in Great Britain*—'is to promote a clear distinction between the standard findings of psycho-analysis and those more speculative accretions which...do more credit to the hearts than to the heads of their sponsors.'

This first volume of Selected Papers on Psychoanalysis is, in virtue of the omissions, as prodigious as the visible peak of an iceberg, and much of it does not make easy reading. It includes 28 papers out of the 177 that Dr Glover has published over a period of forty years, and that are listed in full in a personal bibliography at the end of the book. This selection stretches chronologically from 1942 (*The Significance of the Mouth in Psycho-analysis*) to 1954 (*The Frontiers of Psycho-analysis*), and is amplified, garnished, and enlivened by introductions and notes—expository, hortatory, critical, topical, and controversial—on Eysenck, Goldman-Eisler, the relations between the different disciplines bearing on psychology, the risk of the dilution of psychoanalytic values in the course of inter-disciplinary team-research (not that Glover is opposed to this), and a variety of other matters. Inevitably, the papers are not all of equal significance to the general psychologist, and not all of them, despite the title of the book, and its reiterated emphasis on the value of the developmental approach in clinical practice and theoretical discussion, are equally important to or likely to be understood by the genetic psychologist. But orthodox psychoanalysis, in one of its most significant aspects, is an attempt at a genetic psychological theory, and it is important that the child psychologist should be clearer than he often is about the origin and nature of the formulae that he employs without adequate examination. Yet here, in reading these papers, we are reminded that, whatever his claims, the analyst is in no better case than the experimentalist, like Michotte or Hebb, whose formulations cry out for genetic underpinning. 'The obscure stages of early mental development which are not accessible to direct analysis...can be outlined only by means of plausible "reconstruction"', which exposes the analyst to the dangers of 'encroachment by speculations which, uncontrolled, are little more than wild surmises'. These are Dr Glover's words, and it is this continuing element of 'wild surmise' that still supplies much of the detail of the psychoanalytic account of early psychological development.

But there must be few psychologists, whether they are working in the clinical field or not, who would fail to find stimulation in this collection of essays by a clinician of experience and wisdom. These qualities lend force to the slighter and less technical papers, some of which are among the most interesting in the book; and it is these too, rather than the weight of psychoanalytic theory, that give point to many of Dr Glover's *obiter dicta* as, for example (from *The Unconscious Functions of Education*): 'even amongst analysts the view that education is an unconscious mechanism...serving amongst other purposes the purpose of unconscious instinct inhibition is rarely emphasised', or again from a critical notice (*Psychology and the Public*) written in 1945: 'since defective children are almost invariably handled with unwisdom, they naturally respond by retreating into their lonely interiors, from which it takes a good deal of love and reassurance to entice and rescue them.' The present writer holds the view that there is urgent need for a common understanding, not merely between psychiatrists and professional psychologists 'on the job' but at the level of principle between education, psychology, and medicine (paediatrics, psychiatry, and psychoanalysis), and there is much support for that broad proposition in this record of the development of Dr Glover's thinking over many years of practice, reflexion, and astonishing literary productivity.

GEORGE SETH

CORRIGENDUM. In the review of *Interpersonal Diagnosis of Personality*, the last sentence of the penultimate paragraph on p. 231 should read: 'The Kaiser Foundation Research may well have a like influence on clinical and social studies during the next decade.'